A BERNOULLI TWO-ARMED BANDIT

by

Donald A. Berry

Technical Report No. 151

February, 1971

University of Minnesota
Minneapolis, Minnesota

A BERNOULLI TWO-ARMED BANDIT

SUMMARY

One of two independent Bernoulli processes (arms) with unknown expectations ρ and λ is selected and observed at each of n stages. The selection problem is sequential in that the process which is selected at a particular stage is a function of the results of previous selections as well as of prior information about ρ and λ. The variables ρ and λ are assumed to be independent under the (prior) probability distribution. The objective is to maximize the expected number of successes from the n selections. Sufficient conditions for the optimality of selecting one or the other of the arms are given and illustrated for example distributions. The stay-on-a-winner-rule is proved.

ACKNOWLEDGEMENTS

I gratefully acknowledge the assistance and comments of Professors J. B. Kadane, P. Feder, T. Cover, and especially Professor L. J. Savage, who led me through some troubled waters and pushed me through others.

TABLE OF CONTENTS

A BERNOULLI TWO-ARMED BANDIT

1. Introduction and statement of the problem

Let $\mathcal{R}$ and $\mathcal{L}$ denote independent Bernoulli processes with parameters--probabilities of success-- ρ and λ respectively. Call $\mathcal{R}$ the right arm and $\mathcal{L}$ the left arm. An observation on one of the arms is called a pull. A right pull or a left pull is made at each of n stages and the result of the pull at each stage is known before a right or left pull is made at the next stage. The parameters ρ and λ associated with $\mathcal{R}$ and $\mathcal{L}$ are not known precisely but are themselves random variables. The sequences of successes and failures associated with the right and left arms are therefore not sequences of independent Bernoulli trials, but are independent conditional on the unknown quantities ρ and λ, so that pulls on the right and left arms are exchangeable--see, for example, (Feller 1966, Section VII 4)--rather than independent.

Let I_k denote the pattern of information present about $\mathcal{R}$ and $\mathcal{L}$ at stage $k + 1$; that is, after k pulls. The pattern of information or accumulated data, I_k, can always be regarded as a probability distribution on the unknown parameters ρ and λ. I_0 or I is the initial pattern of information, and consists of an initial probability distribution for each of the parameters ρ and λ. Throughout this paper the parameters are assumed to be initially, and therefore also henceforth, statistically independent. The problem is to decide which arm to pull at stage $k + 1$ conditional on the accumulated data I_k; thus, the results of the first k pulls as well as the initial distributions of ρ and λ can affect the decision at stage $k + 1$.

Let $I = (R, L)$ denote a pair of arbitrary initial distributions; $R = R(\rho)$ and $L = L(\lambda)$. A success on the right arm changes R to a new distribution, say σR, and a success on the left arm changes L to σL; a failure on the right arm changes R to φR and a failure on the left arm changes L to φL. Letting $E(\rho|R)$ and $E(\lambda|L)$ represent the expected values of ρ and λ with respect to R and L:

$$E(\rho|R) = \int_0^1 u\,dR(u), \tag{1.1}$$

$$E(\lambda|L) = \int_0^1 v\,dL(v); \tag{1.2}$$

when the notation $E(\rho)$ or $E(\lambda)$ is used, the distribution R or L will be understood. $E(\rho)$ and $E(\lambda)$ are the probabilities of a success on the first pull on $\mathcal{R}$ and $\mathcal{L}$, respectively, conditional on I.

The expected number of successes over the n stages is to be maximized. This objective is a natural interpretation of a gambler's desire to make as much money as possible. It amounts to assuming that the utility of money (or, success) is linear for the gambler.

An arm should not be selected only because the expected probability of success on that arm is greater than it is on the other, since a future success is worth as much as an immediate success and the other arm may offer a reasonable chance of being better in the long run. $\mathcal{R}$ or $\mathcal{L}$ can be pulled at stage $k + 1$ without waste if by pulling that arm the maximum attainable expected number of successes in the remaining $n - k$ pulls is attained, conditional on I_k. A formal framework for deciding which arm should be pulled is developed first. How this framework relates to the selection problem will be explained presently.

Formally, define

$$W_0^{\mathfrak{R}}(I) = 0, \tag{1.3}$$

$$W_0^{\mathfrak{L}}(I) = 0, \tag{1.4}$$

for all $I = (R, L)$. Where $W_n(I) = \max\{W_n^{\mathfrak{R}}(I), W_n^{\mathfrak{L}}(I)\}$ and as usual $\bar{E}(x) = 1 - E(x)$, recursively define

$$W_n^{\mathfrak{R}}(I) = E(\rho) + E(\rho)W_{n-1}(\sigma R, L) + \bar{E}(\rho)W_{n-1}(\varphi R, L) \tag{1.5}$$

and

$$W_n^{\mathfrak{L}}(I) = E(\lambda) + E(\lambda)W_{n-1}(R, \sigma L) + \bar{E}(\lambda)W_{n-1}(R, \varphi L), \tag{1.6}$$

for $n = 1, 2, \ldots$ and for all $I = (R, L)$.

The right arm is said to be <u>optimal</u> whenever $W_{n-k}(I_k) = W_{n-k}^{\mathfrak{R}}(I_k)$ and the left arm whenever $W_{n-k}(I_k) = W_{n-k}^{\mathfrak{L}}(I_k)$. $W_n(I)$ is the expected worth of (expected number of successes provided by) an optimal procedure, one which maximizes the expected number of successes in the remaining pulls. $W_n^{\mathfrak{R}}(I)$ is the expected worth of pulling $\mathfrak{R}$ and using an optimal procedure thereafter, conditional on the result. A success has worth 1 and obtains on pulling $\mathfrak{R}$ with probability $E(\rho)$; if a success obtains on pulling $\mathfrak{R}$, $I_1 = (\sigma R, L)$, and if a failure obtains on pulling $\mathfrak{R}$, $I_1 = (\varphi R, L)$; since the latter has probability $\bar{E}(\rho)$, Equation (1.5) is in keeping with the definitions of $W_n^{\mathfrak{R}}(I)$ and $W_n(I)$. Equation (1.6) is symmetric with (1.5). The function

$$\Delta_n(I) = W_n^{\mathfrak{R}}(I) - W_n^{\mathfrak{L}}(I), \tag{1.7}$$

is therefore, the expected advantage of choosing $\mathcal{R}$ over $\mathcal{L}$ at the first stage. $\mathcal{R}$ is optimal if $\Delta_n(I) \geq 0$ and $\mathcal{L}$ is optimal if $\Delta_n(I) \leq 0$.

The problem treated in this paper is a special case of the <u>two-armed bandit</u> problem. Other two-armed bandit problems differ in various ways: $\mathcal{R}$ and $\mathcal{L}$ are not necessarily independent; the objective function can be different, as when the weight of a success is a function of the stage at which it occurs (discounting the future, for example); the number of pulls can be infinite, in which case the problem would be uninteresting unless the objective specified is nonetheless finite.

The problem described here is set in discrete time. Chernoff (1968) considers a continuous version where $\mathcal{R}$ and $\mathcal{L}$ are time-continuous processes (in particular, independent Wiener processes with unknown means and known variances). $\mathcal{R}$ or $\mathcal{L}$ is observed, payoff accumulates equal to the value of the process, and information about that process accumulates continuously until a switch is made and the other process is observed. Observation continues until some fixed time has elapsed. A less than helpful characteristic of every optimal selection procedure in this version is that almost every switch is accompanied by an uncountable number of switches within every time interval of positive duration which includes the switch.

Quisel (1965) touches on still another variant in which there is a time delay between a pull and getting information from the pull.

A problem related to (but different from) the two-armed bandit treated here is the two-armed bandit with finite memory. See (Yakowitz 1969) for a description of this problem and for additional references.

For applications of two-armed bandit problems, see (Bradt et al. 1956), (Quisel 1965), and (Dubins and Savage 1965, Chapter 12).

2. The initial distributions

In this section the distributions R and L will be written in a more convenient form and particular patterns of information will be considered.

Consider an arm, say the right arm for definiteness. It will prove useful to regard R as having arisen from another distribution, given by the measure $\mu_{\mathfrak{R}}$, and a number $N_{\mathfrak{R}}$ of pulls on the right arm that yielded, say, r successes and $r' = N_{\mathfrak{R}} - r$ failures. The numbers r and r' are allowed to be real and not just positive integers. Since the pulls are exchangeable, only the numbers of successes and failures affect $\mu_{\mathfrak{R}}$, and R can be written

$$(2.1) \qquad \sigma^r \varphi^{r'} \mu_{\mathfrak{R}} ,$$

regardless of the order of the r successes and r' failures. If information about the right arm is regarded as having arisen in this manner, then according to Bayes' theorem,

$$(2.2) \qquad dR(\rho) = \nu(r, r'; \mu_{\mathfrak{R}})\rho^r(1-\rho)^{r'} d\mu_{\mathfrak{R}}(\rho),$$

where

$$(2.3) \qquad \nu(r, r'; \mu_{\mathfrak{R}}) = [\int_0^1 \rho^r(1-\rho)^{r'} d\mu_{\mathfrak{R}}(\rho)]^{-1}.$$

The distribution R can always be written in the form (2.2); one $(r, r'; \mu_{\mathfrak{R}})$ that qualifies is $(0, 0; R)$, where $\nu(0, 0; R) = 1$. $\mu_{\mathfrak{R}}$ can be any positive measure and r and r' any real numbers for which $\nu(r, r'; \mu_{\mathfrak{R}})$ is positive. The set of (r, r') which satisfy this condition will be called the <u>possibility region</u> for $\mu_{\mathfrak{R}}$. If (r, r')

is in the possibility region for $\mu_{\mathfrak{R}}$, then any point $(r + a, r' + b)$ is also in the possibility region for $\mu_{\mathfrak{R}}$ for nonnegative a and b, provided $\mu_{\mathfrak{R}}$ assigns positive measure to the interior of the unit interval. Therefore, the possibility region for any measure which assigns positive measure to the interior of the unit interval is a quadrant of the (r, r') plane (which may be a half plane or the whole plane) defined by $(r_* + a, r_*' + b)$ for some (r_*, r_*') and all positive a and b. This quadrant may be open or closed depending on $\mu_{\mathfrak{R}}$; either half-line, $r = r_*$ for $r' > r_*'$ or $r' = r_*'$ for $r > r_*$, may be included; if the point (r_*, r_*') is included then both of these half-lines are included. Similarly for the left arm,

$$(2.4) \qquad dL(\lambda) = \nu(\mathit{l}, \mathit{l}'; \mu_{\mathfrak{L}})\lambda^{\mathit{l}}(1-\lambda)^{\mathit{l}'}d\mu_{\mathfrak{L}}(\lambda).$$

Points in the interior of the possibility region for $\mu_{\mathfrak{R}}$ will play a special role in Section 6. Such points (r, r') are characterized by

$$(2.5) \qquad \nu(r + \delta r, r' + \delta r'; \mu_{\mathfrak{R}}) < \infty \quad \text{for} \quad |\delta r|, |\delta r'| \leq \epsilon,$$

for some $\epsilon > 0$.

Since the distribution R is determined by r, r', and $\mu_{\mathfrak{R}}$ and the distribution L is determined by l, l', and $\mu_{\mathfrak{L}}$, the initial pattern of information will sometimes be written:

$$(2.6) \qquad I = (R, L) = (r, r', \mu_{\mathfrak{R}}; L),$$

or sometimes,

$$(2.7) \qquad I = (r, r', \mu_{\mathfrak{R}}; \mathit{l}, \mathit{l}', \mu_{\mathfrak{L}}).$$

A success on $\mathfrak{R}$ then yields the pattern:

$$I_1 = (\sigma R, L) = (r + 1, r', \mu_{\mathfrak{R}}; L), \tag{2.8}$$

according to Bayes' theorem. The expected values of ρ or λ with respect to I will sometimes be written $E(\rho | r, r'; \mu_{\mathfrak{R}})$ and $E(\lambda | \ell, \ell'; \mu_{\mathfrak{L}})$. In this notation, for example, $E(\rho | \sigma R) = E(\rho | r + 1, r'; \mu_{\mathfrak{R}})$.

Arbitrary patterns of information will be investigated. However, patterns where there exist r, r', ℓ, and ℓ' for which $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}}$ are of particular interest. Two cases of this type of pattern will be considered in depth, the first in Section 10 and the second in Section 11.

In the first case there exist positive $r, r', \ell, \ell' < \infty$ for which $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \beta$, where

$$d\beta(x) = x^{-1}(1-x)^{-1}dx. \tag{2.9}$$

In this case R and L are beta distributions. If $\mu_{\mathfrak{R}} = \beta$, then $r_* = r_*' = 0$ and the possibility region for β does not include either of the axes, $r = 0$ or $r' = 0$. The conjugate nature of the beta family of distributions is well known--see, for example, (Raiffa and Schlaifer 1961); if R is a beta distribution then so are σR, φR, $\sigma\varphi R$, etc. The expected value of ρ is particularly simple for this case:

$$E(\rho | r, r'; \beta) = \frac{r}{r + r'} = \frac{r}{N_{\mathfrak{R}}}. \tag{2.10}$$

In the second special case $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \tau$ is a two-point measure, concentrating probability 1/2 on each of τ_1 and τ_2, $\tau_1 < \tau_2$, with not both $\tau_1 = 0$ and $\tau_2 = 1$. (For convenience, it is assumed that τ cannot be a one-point measure--in which case $\tau_1 = \tau_2$. Many results concerning τ apply as well to one-point measures, but none are of interest if $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}}$.) If $\tau_1 > 0$ and $\tau_2 < 1$ then $(r_*, r_*') = (-\infty, -\infty)$ and all points in the (r, r') plane are in the possibility region for τ. If $\tau_1 = 0$ and $\tau_2 < 1$ then $(r_*, r_*') = (0, -\infty)$ and all points for which $r \geq r_*$ are possible. If $\tau_1 > 0$ and $\tau_2 = 1$ then $(r_*, r_*') = (-\infty, 0)$ and all points for which $r' \geq r_*'$ are possible. (If the pair $\tau_1 = 0$ and $\tau_2 = 1$ were allowed, then r_* and r_*' would both be zero and the possibility region for τ would consist of only the nonnegative axes.) The expected value of ρ is

$$(2.11) \qquad E(\rho | r, r'; \tau) = \frac{\tau_1^{r+1}(1-\tau_1)^{r'} + \tau_2^{r+1}(1-\tau_2)^{r'}}{\tau_1^r(1-\tau_1^{r'}) + \tau_2^r(1-\tau_2)^{r'}} .$$

If $\mu_{\mathfrak{R}} = \beta$ then (2.5) holds for all points in the possibility region for $\mu_{\mathfrak{R}}$. If $\mu_{\mathfrak{R}} = \tau$ then (2.5) holds for all points in the possibility region for $\mu_{\mathfrak{R}}$ provided $\tau_1 > 0$ and $\tau_2 < 1$.

3. The dynamic programming approach

The two-armed bandit is a typical problem in dynamic programming. A universal method of solution for such problems with a finite number of time stages will be reviewed and employed in this section. This section, which is independent of later sections, presents a method of straightforward calculation that yields the values $W^{\mathcal{R}}_{n-k}(I_k)$ and $W^{\mathcal{L}}_{n-k}(I_k)$ for every k, $k = 0, 1, \ldots, n$. To begin with, according to (1.3) and (1.4), $W_0(I_n) = 0$ for all possible patterns of information I_n. According to the exchangeability of the pulls, these patterns are determined by the possible combinations of the numbers of successes r and failures r' observed on $\mathcal{R}$ in the n stages and the numbers of successes ℓ and failures ℓ' observed on $\mathcal{L}$ in the n stages; r, r', ℓ, and ℓ' are integers and sum to n. There are

$$(3.1) \qquad \binom{n+3}{3} = \frac{(n+1)(n+2)(n+3)}{6}$$

such patterns. It is not clear whether or not each of these patterns could obtain from a reasonable selection procedure; nevertheless, the method to be described requires consideration of every possible pattern I_n.

Equations (1.5) and (1.6) are used to obtain $W^{\mathcal{R}}_1(I_{n-1})$ and $W^{\mathcal{L}}_1(I_{n-1})$ for all possible I_{n-1}, every pattern of information that could arise from I_0 after $n - 1$ pulls. If I_{n-1} is I_0 changed by r, r', ℓ, and ℓ' as defined above, now with $r + r' + \ell + \ell' = n - 1$, then these are equal respectively to $E(\rho|r, r'; R)$ and $E(\lambda|\ell, \ell'; L)$, the expected values of the distributions $\sigma^r\varphi^{r'}R$ and $\sigma^{\ell}\varphi^{\ell'}L$, since $W_0(I_n) = 0$ for all I_n; that is,

$$(3.2)\qquad W_1^{\mathfrak{R}}(I_{n-1}) = \nu(r,\ r';\ R)\int_0^1 \rho^{r+1}(1-\rho)^{r'}dR(\rho),$$

$$(3.3)\qquad W_1^{\mathfrak{L}}(I_{n-1}) = \nu(l,\ l';\ L)\int_0^1 \lambda^{l+1}(1-\lambda)^{l'}dL(\lambda),$$

and $W_1(I_{n-1})$ is the maximum of these two numbers. In the notation of the previous section,

$$(3.4)\qquad \nu(r,\ r';\ R) = E^{-1}(\rho^r(1-\rho)^{r'}),$$

$$(3.5)\qquad \nu(l,\ l';\ L) = E^{-1}(\lambda^l(1-\lambda)^{l'}),$$

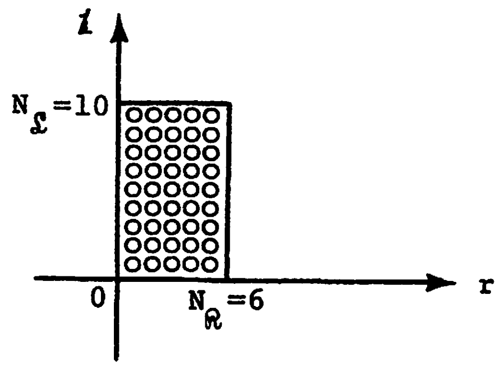
since unconditional expectation is with respect to the pattern $I = (R,\ L)$.

Equations (1.5) and (1.6) can now be used to find $W_2^{\mathfrak{R}}(I_{n-2})$ and $W_2^{\mathfrak{L}}(I_{n-2})$ for all possible I_{n-2}, etc. Proceeding in this manner provides $W_{n-k}(I_k)$ and $\Delta_{n-k}(I_k)$ for each stage k and all patterns of information, I_k, that are possible after k pulls starting with I_0.

As an example, suppose that $n = 12$ and that in equations (2.2) and (2.4), $r = r' = l = l' = 1$ and $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \beta$; that is, R and L are both uniform distributions and are in the beta family; $I_0 = (1,1,\beta;1,1,\beta)$. Since in this example $I = (R,\ L) = (L,\ R)$, it is clear that $\Delta_{12}(I_0) = 0$. Suppose that $\mathfrak{L}$ is pulled and a success observed; $I_1 = (1,1,\beta;2,1,\beta)$. The sign of $\Delta_{11}(I_1)$ is not easy to establish, so that an optimal selection procedure is not obvious. Even less easy to find is the sign of $\Delta_5(1,1,\beta;5,4,\beta)$. The pattern $I_7 = (1,1,\beta;5,4,\beta)$ would occur if $\mathfrak{L}$ were pulled seven times yielding four successes and three failures and $\mathfrak{R}$ not pulled at all. There may not be a reasonable selection procedure which could produce this particular pattern; nonetheless, the values $W_5^{\mathfrak{R}}(1,1,\beta;5,4,\beta)$ and $W_5^{\mathfrak{L}}(1,1,\beta;5,4,\beta)$ are necessary in order to allow the calculation of $W_6(1,1,\beta;5,3,\beta)$, for example, and a reasonable selection procedure may produce the pattern $(1,1,\beta;5,3,\beta)$ without producing $(1,1,\beta;5,4,\beta)$.

For this example, the possible patterns I_{12} corresponding to no remaining pulls are given by $(r,r',\beta;\ell,\ell',\beta)$, where $r = 1,\ldots,13$; $r' = 1,\ldots,14 - r$; $\ell = 1,\ldots,15 - r - r'$; and $\ell' = 16 - r - r' - \ell$.

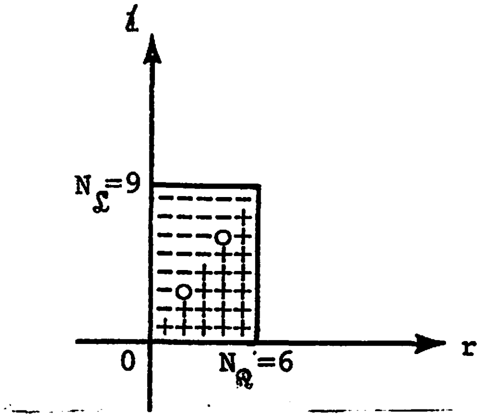
Using the notation $N_{\mathfrak{R}} = r + r'$ and $N_{\mathfrak{L}} = \ell + \ell'$, $N_{\mathfrak{R}} - 2$ and $N_{\mathfrak{L}} - 2$ are the number of times $\mathfrak{R}$ and $\mathfrak{L}$ have been pulled to arrive at $I_{12} = (r,r',\beta;\ell,\ell',\beta)$. It follows that the possible values of $N_{\mathfrak{R}}$ are $2,\ldots, 14$, and $N_{\mathfrak{L}} = 16 - N_{\mathfrak{R}}$. Each possible pattern conditional on $N_{\mathfrak{R}}$ and $N_{\mathfrak{L}}$ is given by a value of r, for $r = 1,\ldots, N_{\mathfrak{R}} - 1$ and a value of ℓ, for $\ell = 1,\ldots, N_{\mathfrak{L}} - 1$. For $N_{\mathfrak{R}} = 6$ and $N_{\mathfrak{L}} = 10$, each such pattern is represented in Figure 3.1 by an "o", indicating that for the corresponding pattern, $\Delta_0(I_{12}) = 0$.



The sign of $\Delta_0(I_{12})$ for $N_{\mathfrak{R}} = 6$, $N_{\mathfrak{L}} = 10$.

FIGURE 3.1

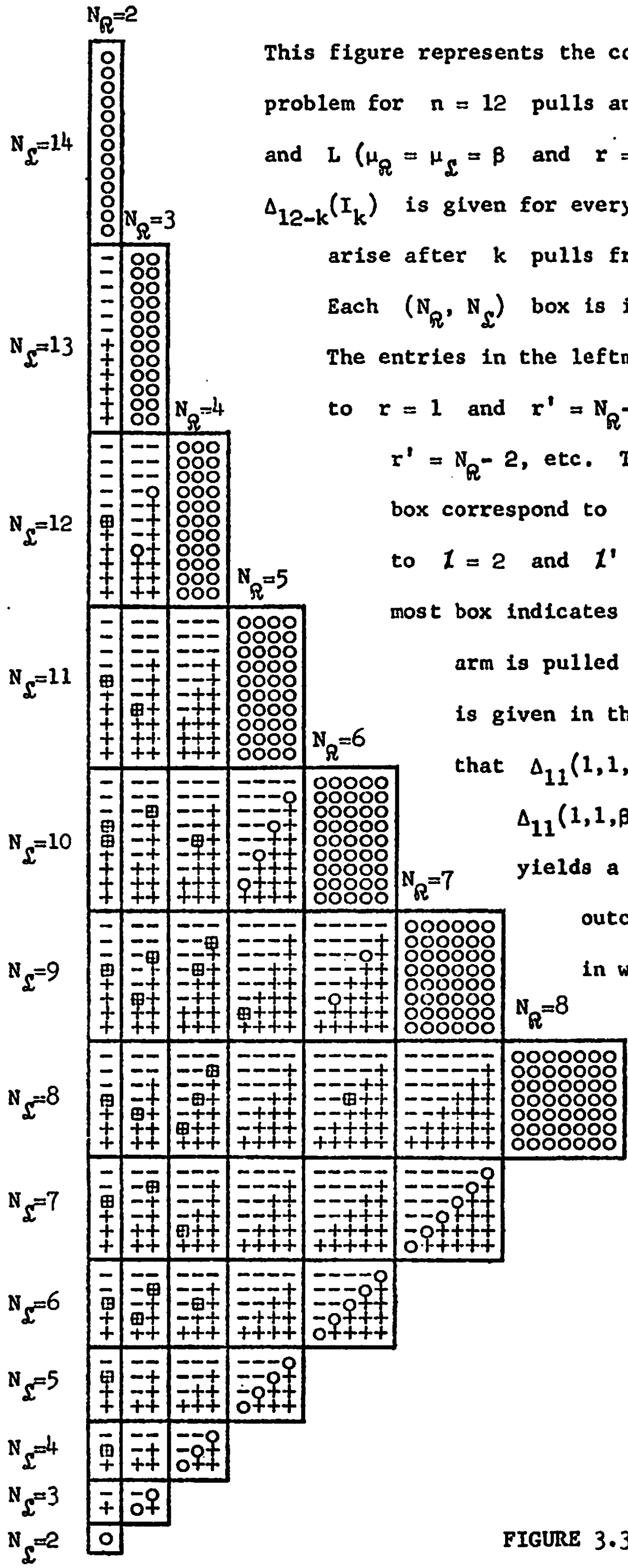
Similarly, the possible patterns after 11 pulls, I_{11}, correspond to one remaining pull and are given by values of $N_{\mathfrak{R}}$, $N_{\mathfrak{L}}$, r, and l, where $N_{\mathfrak{R}} + N_{\mathfrak{L}} = 15$, $r = 1, \ldots, N_{\mathfrak{R}} - 1$; and $\mathit{l} = 1, \ldots, N_{\mathfrak{L}} - 1$. Each possible pattern conditional on $N_{\mathfrak{R}} = 6$ and $N_{\mathfrak{L}} = 9$ is represented in Figure 3.2. The entries in the box of Figure 3.2 indicate the sign of $\Delta_1(I_{11})$ for the corresponding I_{11}; as is evident from the definition of $\Delta_n(I)$, $\Delta_1(I_{11})$ has the same sign as $\frac{r}{N_{\mathfrak{R}}} - \frac{\mathit{l}}{N_{\mathfrak{L}}}$.



The sign of $\Delta_1(I_{11})$ for $N_{\mathfrak{R}} = 6$, $N_{\mathfrak{L}} = 9$.

FIGURE 3.2

The box in Figure 3.1 is duplicated in Figure 3.3; the other boxes in Figure 3.3 for which $N_{\mathfrak{R}} + N_{\mathfrak{L}} = 16$ give the sign of $\Delta_0(I_{12})$ (namely, 0) for all possible patterns that correspond to no remaining pulls. The box in Figure 3.2 is also duplicated in Figure 3.3; the other boxes in Figure 3.3 for which $N_{\mathfrak{R}} + N_{\mathfrak{L}} = 15$ give the sign of $\Delta_1(I_{11})$ for all possible patterns that can occur after 11 pulls and that have $N_{\mathfrak{R}} \leq N_{\mathfrak{L}}$. The patterns with $N_{\mathfrak{R}} > N_{\mathfrak{L}}$ are redundant in view of the symmetry of the problem and are not shown. In general, the boxes



This figure represents the complete solution of a two-armed bandit problem for $n = 12$ pulls and uniform initial distributions R and L ($\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \beta$ and $r = r' = \ell = \ell' = 1$). The sign of $\Delta_{12-k}(I_k)$ is given for every pattern of information which can arise after k pulls from $I_0 = (1,1,\beta;1,1,\beta)$, $k = 0,\ldots, 12$. Each $(N_{\mathcal{R}}, N_{\mathcal{L}})$ box is in the style of Figures 3.1 and 3.2. The entries in the leftmost column in each box correspond to $r = 1$ and $r' = N_{\mathcal{R}} - 1$, the next column to $r = 2$ and $r' = N_{\mathcal{R}} - 2$, etc. The entries in the lowest row in each box correspond to $\ell = 1$ and $\ell' = N_{\mathcal{R}} - 1$, the next row to $\ell = 2$ and $\ell' = N_{\mathcal{R}} - 2$, etc. The "o" in the bottom-most box indicates that $\Delta_{12}(I_0) = 0$. Assume the left arm is pulled producing I_1; the sign of $\Delta_{11}(I_1)$ is given in the $N_{\mathcal{R}} = 2$, $N_{\mathcal{L}} = 3$ box, the "+" indicating that $\Delta_{11}(1,1,\beta;1,2,\beta) > 0$ and the "-" that $\Delta_{11}(1,1,\beta;2,1,\beta) < 0$. Proceeding accordingly yields a complete procedure for all possible outcomes, except that a symmetric part, in which $N_{\mathcal{R}} > N_{\mathcal{L}}$, is not shown. The number of remaining pulls is constant within each $(N_{\mathcal{R}}, N_{\mathcal{L}})$ box and equals $16 - N_{\mathcal{R}} - N_{\mathcal{L}}$. The symbol " ⊞ " is discussed in the text.

FIGURE 3.3

in Figure 3.3 for which $N_{\mathfrak{R}} + N_{\mathfrak{L}} = 4 + k$ give the sign of $\Delta_{12-k}(I_k)$ for all possible patterns I_k, which are given by $r = 1,\ldots, N_{\mathfrak{R}} - 1$ and $\mathcal{l} = 1,\ldots, N_{\mathfrak{L}} - 1$, for $k = 0, 1,\ldots, 12$.

Figures 3.1, 3.2 and 3.3 accentuate the sign of Δ_{n-k}. However, the algorithm described in this section requires the calculation of $W^{\mathfrak{R}}_{n-k}$ and $W^{\mathfrak{L}}_{n-k}$ for each possible pattern, so that the values of Δ_{n-k}, the difference of these two values, could be supplied to the curious.

The patterns $I_k = (r,r',\beta;\mathcal{l},\mathcal{l}',\beta)$ in Figure 3.3 for which $E(\rho|r,r';\beta) = \frac{r}{N_{\mathfrak{R}}} < \frac{\mathcal{l}}{N_{\mathfrak{L}}} = E(\lambda|\mathcal{l},\mathcal{l}';\beta)$, while $\Delta_{12-k}(r,r',\beta;\mathcal{l},\mathcal{l}',\beta) > 0$, deserve special mention and are indicated in Figure 3.3 by "⊞". For these patterns, the probability of success on the initial pull is smaller on arm $\mathfrak{R}$, yet, according to Figure 3.3, the expected value of information to be gained pulling arm $\mathfrak{R}$ makes pulling it worthwhile. Notice that for $N_{\mathfrak{R}} \leq N_{\mathfrak{L}}$ there are no patterns for which simultaneously: $\frac{r}{N_{\mathfrak{R}}} \geq \frac{\mathcal{l}}{N_{\mathfrak{L}}}$ and $\Delta_{12-k}(r,r',\beta;\mathcal{l},\mathcal{l}',\beta) < 0$. This fact, and other apparent regularities in Figure 3.3, will be investigated in later sections.

There are two main drawbacks to the approach described in this section; first, a large number of calculations is necessary; for fixed n the problem is four-dimensional (though my computer program requires only on the order of $n^3/6$ storage locations), and second, the values $\Delta_{n-k}(I_k)$ are found only for patterns I_k on a certain lattice within a four-dimensional simplex for particular measures $\mu_{\mathfrak{R}}$ and $\mu_{\mathfrak{L}}$, and then only for one value of the number of pulls remaining.

4. The function $\Delta_n(I)$

In this section, the function $\Delta_n(I)$ will be defined recursively. From the definition of $\Delta_n(I)$, for all nonnegative n and for any $I = (R, L)$,

$$(4.1) \qquad W_n^{\mathfrak{R}}(I) = W_n(I) + \Delta_n^-(I) ,$$

$$(4.2) \qquad W_n^{\mathfrak{L}}(I) = W_n(I) - \Delta_n^+(I),$$

where, in a slight departure from normal usage, $\Delta_n^-(I) = \min\{0, \Delta_n(I)\}$ and $\Delta_n^+(I) = \max\{0, \Delta_n(I)\}$. In view of (4.2), for $n \geq 1$ (1.5) becomes

$$(4.3) \qquad W_n^{\mathfrak{R}}(I) = E(\rho) + E(\rho)[W_{n-1}^{\mathfrak{L}}(\sigma R, L) + \Delta_{n-1}^+(\sigma R, L)] + \overline{E}(\rho)[W_{n-1}^{\mathfrak{L}}(\varphi R, L) + \Delta_{n-1}^+(\varphi R, L)];$$

and in view of (4.1), (1.6) becomes

$$(4.4) \qquad W_n^{\mathfrak{L}}(I) = E(\lambda) + E(\lambda)[W_{n-1}^{\mathfrak{R}}(R, \sigma L) - \Delta_{n-1}^-(R, \sigma L)] + \overline{E}(\lambda)[W_{n-1}^{\mathfrak{R}}(R, \varphi L) - \Delta_{n-1}^-(R, \varphi L)].$$

For $n \geq 2$, the terms

$$(4.5) \qquad E(\rho) + E(\rho)W_{n-1}^{\mathfrak{L}}(\sigma R, L) + \overline{E}(\rho)W_{n-1}^{\mathfrak{L}}(\varphi R, L),$$

in (4.3) amount to the expected worth of the following procedure: Pull $\mathfrak{R}$ first and $\mathfrak{L}$ second, and use an optimal procedure thereafter. (Of course there can be no second pull if $n < 2$.) Likewise,

$$(4.6) \qquad E(\lambda) + E(\lambda)W_{n-1}^{\mathfrak{R}}(R, \sigma L) + \overline{E}(\lambda)W_{n-1}^{\mathfrak{R}}(R, \varphi L)$$

is the expected worth of the procedure: Pull $\mathfrak{L}$ first and $\mathfrak{R}$ second, and use an optimal procedure thereafter. Since the pulls are exchangeable, this interpretation means that the expressions (4.5) and (4.6) are equal for $n \geq 2$. Therefore, subtracting (4.4) from (4.3) yields for $n \geq 2$,

$$\Delta_n(I) = E(\rho)\Delta^+_{n-1}(\sigma R, L) + \overline{E}(\rho)\Delta^+_{n-1}(\varphi R, L) + E(\lambda)\Delta^-_{n-1}(R, \sigma L) + \overline{E}(\lambda)\Delta^-_{n-1}(R, \varphi L). \tag{4.7}$$

This is a promising expression for $\Delta_n(I)$ since, together with the evident initial condition,

$$\Delta_1(I) = E(\rho) - E(\lambda), \tag{4.8}$$

(4.7) defines $\Delta_n(I)$ recursively.

It seems reasonable to expect that the vanishing of particular terms in (4.7) implies the vanishing of other terms. It will be shown that $\Delta_{n-1}(\sigma R, L) > 0$ whenever $\Delta_{n-1}(\varphi R, L) > 0$ and symmetrically, $\Delta_{n-1}(R, \sigma L) < 0$ whenever $\Delta_{n-1}(R, \varphi L) < 0$, in Section 5 for $n = 2$ (Theorems 5.1 and 5.2) and in Section 8 for general n (Theorem 8.1).

Several facts about the function Δ_n are easy to verify. Three intuitive theorems will be proved formally.

Since Δ_n is the expected advantage of choosing $\mathfrak{R}$ over $\mathfrak{L}$, clearly,

$$-1 \leq \Delta_n \leq 1; \tag{4.9}$$

in fact, more can be said.

Theorem 4.1.

For any pattern of information $I = (R, L)$,

$$-\overline{E}(\rho) \leq \Delta_n(I) \leq \overline{E}(\lambda) \tag{4.10}$$

for all n.

Proof:

The conclusion is obvious for $n = 1$ according to (4.8). If (4.10) holds at $n - 1$, then according to (4.7),

$$\Delta_n(I) \geq E(\lambda)[-\overline{E}(\rho)] + \overline{E}(\lambda)[-\overline{E}(\rho)] = -\overline{E}(\rho), \tag{4.11}$$

$$\Delta_n(I) \leq E(\rho)E(\lambda) + \overline{E}(\rho)\overline{E}(\lambda) = \overline{E}(\lambda). \ \square \tag{4.12}$$

If either of the distributions R or L is of a particular type, $\Delta_n(I)$ may be easy to calculate for all n. For example, if an arm yields success with probability one, then it should be pulled, and the expected loss due to pulling the other arm is the difference between 1 and its expected worth.

Theorem 4.2.

If R is a one-point distribution, concentrating probability one at $\rho = 1$, then for all n and L,

$$\Delta_n(I) = \overline{E}(\lambda), \tag{4.13}$$

which is nonnegative.

Proof:

For $n = 1$, (4.13) is implied by (4.8) since $E(\rho) = 1$. Assuming the result at $n - 1$, (4.7) implies

(4.14) $\Delta_n(I) = \Delta_{n-1}(\sigma R, L) = \bar{E}(\lambda)$. □

In general, the worth of pulling a particular arm consists in the net worth with respect to expected immediate payoff and with respect to worth of information, so that Δ_n is seldom given by the difference in expected immediate payoff,

(4.15) $E(\rho) - E(\lambda)$.

Of course, as seen in (4.8), (4.15) gives this difference when only one pull remains, for then any information gained on the pull will not be used and therefore has no value. For $n \geq 2$, Δ_n is given by (4.15) when and only when the worth of infomation is the same for both arms. This can happen when, for example, (a) both arms are the same, (b) pulling neither arm has information value, or (c) pulling either arm once will give complete information. The next theorem treats these three special patterns of information.

Theorem 4.3.

If I is such that either

(a) $R = L$; that is, the arms are identical initially,

(b) R and L are one-point distributions, concentrating probability one on $E(\rho)$ and $E(\lambda)$; that is, the probability of success is known for both arms,

(c) R and L are two-point distributions, concentrating all the probability at 0 and 1, so that $\rho = 1$ and $\lambda = 1$ with probabilities $E(\rho)$ and $E(\lambda)$; that is, each arm will yield either all successes or all failures and one pull on either arm determines the quality of that arm, then, for $n \geq 1$,

$$\Delta_n(I) = E(\rho) - E(\lambda). \tag{4.16}$$

Theorem 4.3 will be proved algebraically to illustrate notation and the use of (4.7), but each result holds for an intuitive reason that can be made rigorous. The conclusion is obvious in case (a), since $E(\rho) = E(\lambda)$ and $W_n^{\mathcal{R}} = W_n^{\mathcal{L}}$ when $R = L$. In case (b) the quality of both arms is known, and any pull on the inferior arm costs the difference in the quality of the arms. In case (c), the better arm to pull (if indeed one is better than the other) becomes known immediately after the first pull, whichever arm is pulled first (and will yield either all successes, or all failures if both ρ and λ are 0), therefore the difference between pulling the right and left arm is simply the difference in the expected immediate payoffs.

<u>Proof of Theorem 4.3</u>:

Equation (4.16) holds at $n = 1$ for all I. Assuming (4.16) at $n - 1$ for the pattern of information defined in (a), (b), or (c) above, (4.7) implies

$$\begin{aligned} \Delta_n(I) &= \max\{E(\rho), E(\lambda)\}(E(\rho) - E(\lambda)) \\ &\quad + (1 - \max\{E(\rho), E(\lambda)\})(E(\rho) - E(\lambda)) \\ &= E(\rho) - E(\lambda). \ \square \end{aligned} \tag{4.17}$$

5. An example: $n = 2$

By way of example, the function $\Delta_2(I)$ will be explored in this section. Letting unconditional E continue to denote expectation with respect to I, which is the pair of initial distributions R and L,

$$(5.1) \qquad E(f(\rho)) = \int_0^1 f(\rho)\,dR(\rho),$$

$$(5.2) \qquad E(g(\lambda)) = \int_0^1 g(\lambda)\,dL(\lambda);$$

according to (4.7) and in view of (4.8),

$$(5.3) \qquad \Delta_2(I) = E(\rho)\left[\frac{E(\rho^2)}{E(\rho)} - E(\lambda)\right]^+ + \bar{E}(\rho)\left[\frac{E(\rho-\rho^2)}{E(1-\rho)} - E(\lambda)\right]^+$$

$$+ E(\lambda)\left[E(\rho) - \frac{E(\lambda^2)}{E(\lambda)}\right]^- + \bar{E}(\lambda)\left[E(\rho) - \frac{E(\lambda-\lambda^2)}{E(1-\lambda)}\right]^-$$

$$= \left[E(\rho^2) - E(\rho)E(\lambda)\right]^+ + \left[E(\rho) - E(\rho^2) - E(\lambda) + E(\rho)E(\lambda)\right]^+$$

$$+ \left[E(\rho)E(\lambda) - E(\lambda^2)\right]^- + \left[E(\rho) - E(\rho)E(\lambda) - E(\lambda) + E(\lambda^2)\right]^- .$$

For each of the four possible I_1, $\Delta_1(I_1)$ can be positive or negative, so that there are sixteen candidates for the form of $\Delta_2(I)$. However, according to the next three theorems, only eight of the forms are possible.

Theorem 5.1.

For all I, $\Delta_1(\sigma R, L) \geq \Delta_1(\varphi R, L)$, with equality if and only if R is a one-point distribution.

Proof:

For convenience, let

$$(5.4) \qquad \rho^* = \frac{E(\rho^2)}{E(\rho)}, \quad \rho_* = \frac{E(\rho) - E(\rho^2)}{1 - E(\rho)},$$

$$(5.5) \qquad \lambda^* = \frac{E(\lambda^2)}{E(\lambda)}, \quad \lambda_* = \frac{E(\lambda) - E(\lambda^2)}{1 - E(\lambda)}.$$

The theorem says that $\rho^* \geq \rho_*$, with equality if and only if R is a one-point distribution. This is a straightforward consequence of the well-known moment inequality:

$$(5.6) \qquad E(\rho^2) \geq E^2(\rho),$$

and the fact that $E(\rho^2) = E^2(\rho)$ if and only if R is a one-point distribution.

Inequality (5.6) is proved using the linearity of E and the fact that $[\rho - E(\rho)]^2$ is nonnegative. The proof of (5.6) is simple and mentioned here only because a more complicated proof of the same variety occurs later (Lemma 6.7).

Applying (5.6) twice,

$$(5.7) \qquad \rho^* = \frac{E(\rho^2)}{E(\rho)} \geq \frac{E^2(\rho)}{E(\rho)} = E(\rho) = \frac{E(\rho) - E^2(\rho)}{1 - E(\rho)} \geq \frac{E(\rho) - E(\rho^2)}{1 - E(\rho)} = \rho_*.$$

Both inequalities in (5.7) are strict unless R is a one-point distribution. □

<u>Theorem 5.2.</u>

For all I, $\Delta_1(R, \sigma L) \leq \Delta_1(R, \varphi L)$, with equality if and only if L is a one-point distribution.

<u>Proof:</u>

With the roles of $\mathfrak{R}$ and $\mathfrak{L}$ reversed, Theorem 5.2 is seen to be a special case of Theorem 5.1. □

<u>Theorem 5.3</u>.

For all I, at least one of the four terms on the right-hand side of (5.3) is zero.

<u>Proof</u>:

In view of Theorems 5.1 and 5.2, none of the four terms in (5.3) vanish if and only if both of the following hold:

$$\rho_* > E(\lambda), \tag{5.8}$$

$$\lambda_* > E(\rho). \tag{5.9}$$

According to (5.6), (5.8) implies that $E(\rho) > E(\lambda)$ and (5.9) implies that $E(\rho) < E(\lambda)$; the conclusion follows by contradiction. □

Theorems 5.1 and 5.2 remove the need for considering seven of the sixteen candidate forms of $\Delta_2(I)$, and Theorem 5.3 removes another. The form of $\Delta_2(I)$ for the remaining eight cases is shown below; each case is indexed by an ordered quadruple which indicates the sign of each of the terms of (5.3) corresponding to that case:

(5.10) $\Delta_2(I) =$

(+000)	$E(\rho^2) - E(\rho)E(\lambda)$	if $\rho^* > E(\lambda) \geq \rho_*$ and $E(\rho) \geq \lambda^*$,
(++00)	$E(\rho) - E(\lambda)$	if $\rho_* > E(\lambda)$ and $E(\rho) \geq \lambda^*$,
(++-0)	$E(\rho) - E(\lambda) - E(\lambda^2) + E(\rho)E(\lambda)$	if $\rho_* > E(\lambda)$ and $\lambda^* > E(\rho)$,

(+0-0)	$E(\rho^2) - E(\lambda^2)$	if $\rho^* > E(\lambda) \geq \rho_*$ and $\lambda^* > E(\rho) \geq \lambda_*$,
(+0--)	$E(\rho) - E(\lambda) + E(\rho^2) - E(\rho)E(\lambda)$	if $\rho^* > E(\lambda) \geq \rho_*$ and $\lambda_* > E(\rho)$,
(00--)	$E(\rho) - E(\lambda)$	if $E(\lambda) \geq \rho_*$ and $\lambda_* > E(\rho)$,
(00-0)	$E(\rho)E(\lambda) - E(\lambda^2)$	if $E(\lambda) \geq \rho^*$ and $\lambda^* > E(\rho) \geq \lambda_*$,
(0000)	0	if $E(\lambda) \geq \rho^*$ and $E(\rho) \geq \lambda^*$.

Case (0000) in (5.10) corresponds to a very special pattern of information.

Theorem 5.4.

If all four terms on the right hand side of equation (5.3) vanish (in which case $\Delta_2(I) = 0$), then R and L are the same one-point distribution; that is,

$$E(\rho^2) = E^2(\rho) = E^2(\lambda) = E(\lambda^2). \tag{5.11}$$

Proof:

In view of Theorems 5.1 and 5.2, all four terms in (5.3) vanish if and only if,

$$E(\rho)E(\lambda) \geq E(\rho^2) \tag{5.12}$$

and

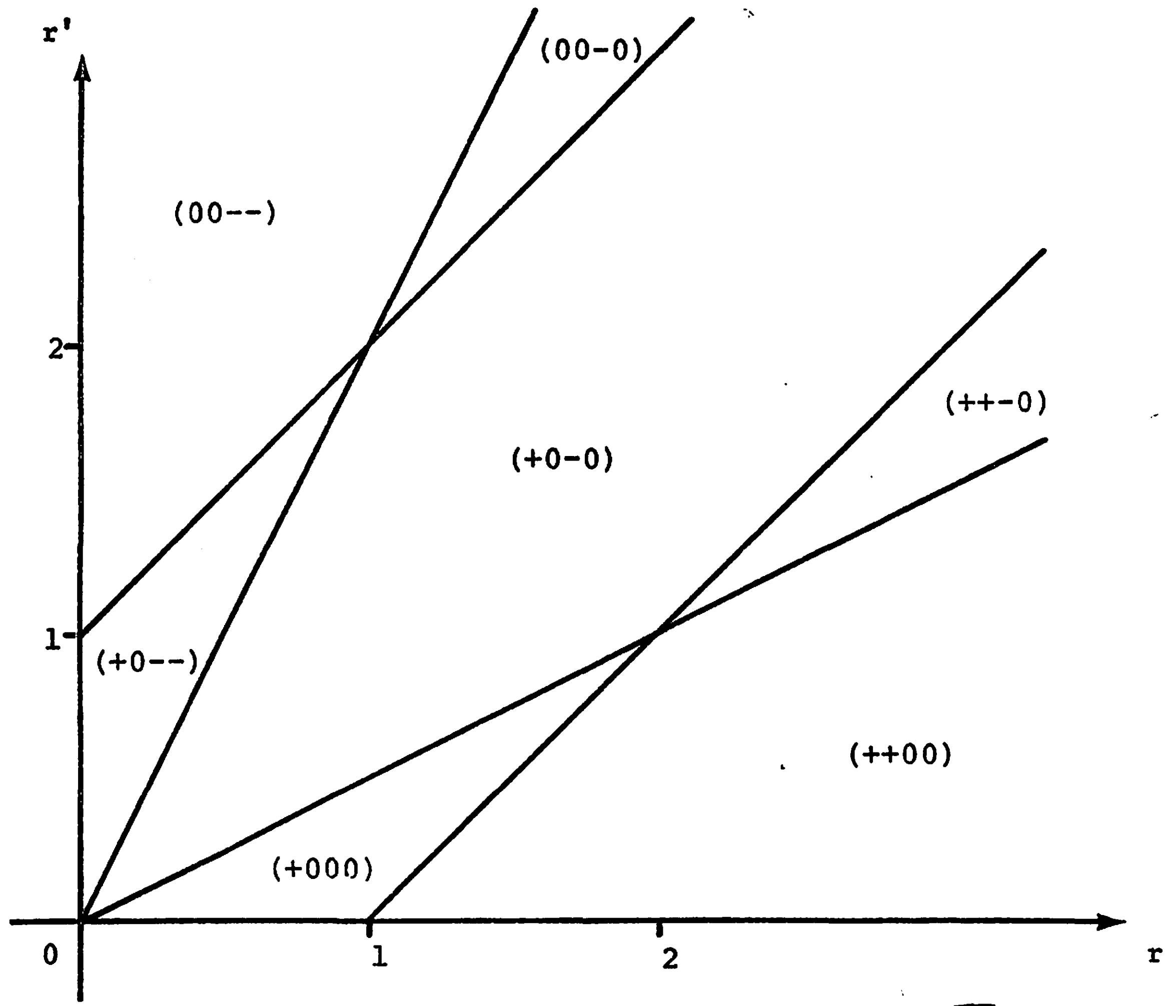
$$E(\rho)E(\lambda) \geq E(\lambda^2), \tag{5.13}$$

which are the conditions given in (5.10), case (0000). According to (5.6), (5.12) implies that $E(\rho) \leq E(\lambda)$ and (5.13) implies that $E(\rho) \geq E(\lambda)$; therefore, $E(\rho) = E(\lambda)$, and equality holds in (5.12) and (5.13). □

In later sections, the measure $\mu_{\mathfrak{R}}$ and the distribution L will be fixed and $\Delta_n(r,r',\mu_{\mathfrak{R}}; L)$ examined as a function of (r, r').

Figure 5.1 shows which regions of a particular (r, r') plane correspond to the eight forms of $\Delta_2(I)$; for the purpose of the figure, $\mu_{\mathfrak{R}} = \beta$, which is defined in (2.9). Such a plane is determined by a distribution L (or by an ℓ and ℓ' for $\mu_{\mathfrak{L}}$ fixed), which in Figure 5.1 is the uniform distribution (for example, $\mu_{\mathfrak{L}}$ could be β and $\ell = \ell' = 1$). Theorem 5.4 says that case (0000) does not apply in any region of the plane in Figure 5.1 (or even in a region of the extended plane) since L is not a one-point distribution .

Theorems 5.1 and 5.2 of this section are trivial consequences of results in later sections. The present section is presented primarily as a didactic orientation for those sections.



The form of $\Delta_2(r,r',\beta;l,l',\beta)$ in the (r, r') plane defined by $l = l' = 1$; the quadruples refer to the cases of (5.10).

FIGURE 5.1

6. Fundamental inequalities

Thus far the possibility that r, r', l, and l' are real numbers and not necessarily integers has not been exploited. This section exploits, and is largely based on, the continuity of $\Delta_n(I)$ in (r, r') in the interior of the possibility region for $\mu_{\mathfrak{R}}$.

Inequalities in $I = (r, r', \mu_{\mathfrak{R}}; L)$ for the function $\Delta_n(I)$ are derived in this section when (r, r') is an interior point of the possibility region for $\mu_{\mathfrak{R}}$. These inequalities will be strengthened in Section 7 and extended to all points in the possibility region. This separate treatment eliminates the need for considering in this section distributions which would unnecessarily complicate the presentation of the basic theory. Results are stated and derived in terms of the right arm; symmetric results hold for the left arm as applications of those for the right arm, with names reversed.

For the purposes of this section, write $I = (R, L)$ as $(r, r', \mu_{\mathfrak{R}}; L)$, where R is given by (2.2) and is a probability distribution for (r, r') in the possibility region for $\mu_{\mathfrak{R}}$.

As will be seen, some important properties of $E(\rho | r, r'; \mu_{\mathfrak{R}})$ are passed on to $\Delta_n(r, r', \mu_{\mathfrak{R}}; L)$ for all n. This motivates studying the behavior of the function

$$(6.1) \qquad \nu(r, r'; \mu_{\mathfrak{R}}) = \int_0^1 \rho^r (1-\rho)^{r'} d\mu_{\mathfrak{R}}(\rho),$$

which will sometimes be abbreviated to $\nu(r, r')$.

Lemma 6.1.

For (r, r') in the interior of the possibility region for $\mu_{\mathfrak{R}}$, there is a positive ε such that for $|\delta r|, |\delta r'| \leq \varepsilon$,

$$(6.2)\qquad \nu(r+\delta r,\ r'+\delta r') = \sum_{0\le s,t} \frac{1}{s!t!}\left[\int_0^1 (\log \rho)^s(\log(1-\rho))^t\rho^r(1-\rho)^{r'}d\mu_{\mathfrak{R}}(\rho)\right] \cdot (\delta r)^s(\delta r')^t,$$

where the series is absolutely convergent. Also, for all $s,\ t \ge 0$,

$$(6.3)\qquad \frac{\partial^{s+t}}{\partial r^s \partial r'^t}\,\nu(r,\ r') = \int_0^1 (\log \rho)^s(\log(1-\rho))^t\rho^r(1-\rho)^{r'}d\mu_{\mathfrak{R}}(\rho).$$

Remark.

For many readers the asserted analyticity of ν in the pair $(r,\ r')$ will be familiar, but it seems easier to give a demonstration than to provide an exactly appropriate reference.

Proof of Lemma 6.1:

For all $\delta r,\ \delta r'$,

$$(6.4)\qquad \rho^{r+\delta r}(1-\rho)^{r'+\delta r'} = \rho^r(1-\rho)^{r'}\exp(\delta r \log \rho)\exp(\delta r' \log(1-\rho))$$
$$= \rho^r(1-\rho)^{r'}\sum_{0\le s,t}\frac{1}{s!t!}(\log \rho)^s(\log(1-\rho))^t(\delta r)^s(\delta r')^t,$$

is absolutely convergent and the partial sums of the series in (6.4) are majorized in absolute value by $\rho^{r-|\delta r|}(1-\rho)^{r'-|\delta r'|}$. Since $\nu(r-\epsilon,\ r'-\epsilon) < \infty$ for sufficiently small ϵ, the Lebesgue dominated convergence theorem applies to (6.4) to prove (6.2).

Repeated differentiation of the convergent power series (6.2) yields (6.3). □

Lemma 6.2.

For all n and $I = (r,\ r',\ \mu_{\mathfrak{R}};\ \mathbb{L})$, $\Delta_n(I)$ is continuous in $(r,\ r')$ in the interior of the possibility region for $\mu_{\mathfrak{R}}$.

Proof:

According to (4.8) and (6.1), the definition of ν,

$$\Delta_1(I) = \frac{\nu(r+1, r')}{\nu(r, r')} - E(\lambda), \tag{6.5}$$

which is continuous in (r, r') in view of Lemma 6.1 for (r, r') in the interior of the possibility region for $\mu_{\Re}$.

If any function g is continuous so are g^+ and g^-. Therefore $\Delta_n(I)$ is continuous by induction, since according to (4.7), it is the sum of four continuous functions. □

Though continuous, $\Delta_n(I)$ is not necessarily everywhere differentiable. Yet $\Delta_n(I)$ is regular in (r, r') except along certain curves (for example, the lines in Figure 5.1); however, the regularity of $\Delta_n(I)$ will not here be analyzed beyond the extent essential for later demonstrations.

The directional derivative of a function $g(x, x')$ along the vector (a, b) is defined to be:

$$D_{(a,b)}g(x, x') = \lim_{h \downarrow 0} \frac{g(x+ha, x'+hb) - g(x, x')}{h} \tag{6.6}$$

(where $h\downarrow 0$ indicates that the limit is taken as h approaches 0 from above), provided the limit in (6.6) exists. $D_{(a,b)}g$ reflects the gradient of g and is the rate of change of g for a point leaving (x, x') with velocity (a, b).

The directional derivative of $g(x, x')$ for the vector (a, b) may exist at a point though neither partial derivative, $\frac{\partial}{\partial x}$ or $\frac{\partial}{\partial x'}$, of $g(x, x')$ exists at the point. However, the directional derivative is a linear function of the partial derivatives provided they do exist and are continuous.

Lemma 6.3.

If both partial derivatives of $g(x, x')$ are continuous at a point, then, at that point,

$$(6.7) \qquad D_{(a,b)}g(x, x') = (a \frac{\partial}{\partial x} + b \frac{\partial}{\partial x'})g(x, x').$$

For the proof of Lemma 6.3, see any advanced calculus text, for example, (Widder 1961, Theorem 9, p. 40).

The regularity of the function $\Delta_n(r, r', \mu_{\mathfrak{R}}; L)$ required for later demonstrations is assured by the next lemma.

Lemma 6.4.

For all n and $I = (r, r', \mu_{\mathfrak{R}}; L)$, the directional derivative $D_{(a,b)}\Delta_n(I)$ exists along every vector (a, b), at every point (r, r') in the interior of the possibility region for $\mu_{\mathfrak{R}}$.

Proof:

For ν defined by (6.1), it is clear from (6.2) that

$$(6.8) \qquad \nu(r + \delta r, r' + \delta r') = \nu(r, r') + \frac{\partial}{\partial r} \nu(r, r')\delta r + \frac{\partial}{\partial r'} \nu(r, r')\delta r' + o(|\delta r| + |\delta r'|),$$

where as usual,

$$(6.9) \qquad \lim_{x \to 0} \frac{o(x)}{x} = 0.$$

In view of (4.8) and (6.1),

$$(6.10) \qquad D_{(a,b)}\Delta_1(I) = D_{(a,b)}E(\rho|r, r'; \mu_{\mathfrak{R}}) = D_{(a,b)} \frac{\nu(r+1, r')}{\nu(r, r')} ;$$

therefore,

$$(6.11) \qquad D_{(a,b)}\Delta_1(I) = \left(a \frac{\partial}{\partial r} + b \frac{\partial}{\partial r'}\right) \frac{\nu(r+1, r')}{\nu(r, r')}$$

according to Lemma 6.1, which proves the lemma for $n = 1$ in view of Lemma 6.3.

If the directional derivative of g exists in any direction (a, b), then the directional derivatives of g^+ and g^- also exist in that direction. The conclusion follows immediately by induction in view of (5.7). □

(It is clear from the proof of Lemma 6.4 that the directional derivative of $\Delta_n(I)$ can be expressed linearly in terms of its partial derivatives except possibly at points where $\Delta_{n-k}(I_k) = 0$ for some pattern of information I_k that can occur after k pulls when starting from I.)

Lemmas 6.2 and 6.4 will be used to prove Theorem 6.1, which will then be extended by Theorems 7.1, 7.2 and 7.3. It seems reasonable to expect that when r is increased, the advantage of pulling $\mathfrak{R}$ over $\mathfrak{L}$ does not decrease, for then $\mathfrak{R}$ promises to be at least as successful as before. Correspondingly, if r' is increased, the advantage of pulling $\mathfrak{R}$ over $\mathfrak{L}$ ought not increase, for then $\mathfrak{R}$ promises to be no more successful than before.

The next theorem says this and more when (r, r') is an interior point of the possibility region for $\mu_{\mathfrak{R}}$: if r and r' increase simultaneously, the advantage of pulling $\mathfrak{R}$ over $\mathfrak{L}$ does not decrease if the rate of change of r compared with the rate of change of r' is larger than a particular bound and does not increase if this ratio is smaller than another (obviously, smaller) bound. These bounds are implied by the following two statements, which put the propositions $J(n)$ and $K(n)$ of the theorem into words. If the probability of a success on $\mathfrak{R}$ conditional on having already observed $n - 1$ successes in $n - 1$ pulls on $\mathfrak{R}$ (this probability is given by the expected value of ρ with respect to $\sigma^{n-1}R$) does not decrease for a particular direction from a particular point in the (r, r') plane, then Δ_n at that point does not decrease for the same direction. If the probability of a failure on $\mathfrak{R}$ conditionally on having already observed $n - 1$ failures in $n - 1$ pulls on $\mathfrak{R}$ (given by the expected value of $1 - \rho$ with respect to $\varphi^{n-1}R$) does not increase for a particular direction from a particular point in the (r, r') plane, then Δ_n at that point does not increase for the same direction.

Theorem 6.1.

For fixed $\epsilon > 0$, provided

$$(6.12) \qquad \nu(r + \delta r, r' + \delta r') < \infty \quad \text{for} \quad |\delta r|, |\delta r'| \leq \epsilon,$$

the following statements are true for $n \geq 1$, for $I = (r, r', \mu_{\mathfrak{R}}; L)$, and for a and b nonnegative and not both 0:

$$J(n): \quad D_{(a,b)}\Delta_n(I) \geq 0 \quad \text{if} \quad D_{(a,b)}E(\rho|r+n-1, r'; \mu_{\mathfrak{R}}) \geq 0,$$

$$K(n): \quad D_{(a,b)}\Delta_n(I) \leq 0 \quad \text{if} \quad D_{(a,b)}E(\rho|r, r'+n-1; \mu_{\mathfrak{R}}) \leq 0.$$

For the proof of Theorem 6.1, which will be presented gradually, the behavior of $E(\rho|r, r'; \mu_{\mathfrak{R}})$ in (r, r') will be needed. Though for $n \geq 2$ the partial derivatives of $\Delta_n(r, r', \mu_{\mathfrak{R}}; L)$ with respect to r and r' do not always exist, the partial derivatives of

$E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$ and $E(\rho|r, r'+n-1; \mu_{\mathfrak{R}})$ do exist and are continuous. These facts are cited in the proof of Lemma 6.4 using Lemma 6.1; they are recorded here for completeness.

Lemma 6.5.

The partial derivatives, $\frac{\partial}{\partial r} E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$, $\frac{\partial}{\partial r'} E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$, $\frac{\partial}{\partial r} E(\rho|r, r'+n-1; \mu_{\mathfrak{R}})$, and $\frac{\partial}{\partial r'} E(\rho|r, r'+n-1; \mu_{\mathfrak{R}})$ exist and are continuous in (r, r') in the interior of the possibility region for $\mu_{\mathfrak{R}}$.

In view of Lemmas 6.3 and 6.5, for $n \geq 1$ the latter directional derivative in $J(n)$ and $K(n)$ of Theorem 6.1 can be written as in (6.7), making these hypotheses easier to manipulate: for $\mu_{\mathfrak{R}}$ not a one-point measure,

$$J(n): \quad D_{(a,b)}\Delta_n(I) \geq 0 \quad \text{if} \quad \frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}});$$

$$K(n): \quad D_{(a,b)}\Delta_n(I) \leq 0 \quad \text{if} \quad \frac{a}{b} \leq A(r, r'+n-1; \mu_{\mathfrak{R}});$$

where

$$(6.13) \qquad A(r, r'; \mu_{\mathfrak{R}}) = \frac{-\frac{\partial}{\partial r'} E(\rho|r, r'; \mu_{\mathfrak{R}})}{\frac{\partial}{\partial r} E(\rho|r, r'; \mu_{\mathfrak{R}})} .$$

The function $A(r, r'; \mu_{\mathfrak{R}})$ would not be defined for one-point measures $\mu_{\mathfrak{R}}$.

Because they are simple, and therefore potentially helpful for following later arguments, the versions of $J(n)$ and $K(n)$ for two special cases will now be given.

First, where $\beta(\rho)$ is defined by (2.9),

$$(6.14) \qquad A(r, r'; \beta) = \frac{r}{r'} \qquad \text{for} \quad r, r' > 0.$$

If $I = (r, r', \beta; L)$, $J(n)$ and $K(n)$ of Theorem 6.1 become

$$J_\beta(n): \quad D_{(a,b)}\Delta_n(I) \geq 0 \quad \text{if} \quad \frac{a}{b} \geq \frac{r + n - 1}{r'},$$

$$K_\beta(n): \quad D_{(a,b)}\Delta_n(I) \leq 0 \quad \text{if} \quad \frac{a}{b} \leq \frac{r}{r' + n - 1}.$$

Second, where R is an interior two-point distribution (so that $\mu_{\mathcal{R}} = \tau$) $J(n)$ and $K(n)$ together completely determine the gradient of $\Delta_n(r, r', \tau; L)$ in (r, r'). If τ concentrates mass $\frac{1}{2}$ on both $\rho = \tau_1$ and on $\rho = \tau_2$, $0 < \tau_1 < \tau_2 < 1$, then,

$$A(r, r'; \tau) = A(\tau) = \frac{\log \dfrac{1 - \tau_1}{1 - \tau_2}}{\log \dfrac{\tau_1}{\tau_2}}, \tag{6.15}$$

which does not depend on r or r'. If $I = (r, r', \tau; L)$, $J(n)$ and $K(n)$ of Theorem 6.1 become

$$J_\tau(n): \quad D_{(a,b)}\Delta_n(I) \geq 0 \quad \text{if} \quad \frac{a}{b} \geq A(\tau),$$

$$K_\tau(n): \quad D_{(a,b)}\Delta_n(I) \leq 0 \quad \text{if} \quad \frac{a}{b} \leq A(\tau).$$

The proof of Theorem 6.1 will depend on the behavior of $A(r, r'; \mu_{\mathcal{R}})$. Using the notation:

$$\mathrm{Cov}(U, V) = E(UV) - E(U)E(V),$$

for real U and V on $[0, 1]$, where unconditional expectation E is as usual with respect to $I = (R, L)$,

$$\frac{\partial}{\partial r} E(\rho \mid r, r'; \mu_{\mathcal{R}}) = \mathrm{Cov}(\rho, \log \rho), \tag{6.17}$$

$$\frac{\partial}{\partial r'} E(\rho \mid r, r'; \mu_{\mathcal{R}}) = \mathrm{Cov}(\rho, \log(1-\rho)), \tag{6.18}$$

in view of (6.3). Therefore,

$$(6.19) \qquad A(r, r'; \mu_{\mathfrak{R}}) = \frac{-\mathrm{Cov}(\rho, \log(1-\rho))}{\mathrm{Cov}(\rho, \log \rho)} = \frac{\mathrm{Cov}(1-\rho, \log(1-\rho))}{\mathrm{Cov}(\rho, \log \rho)} .$$

<u>Lemma 6.6.</u>

Provided $\mu_{\mathfrak{R}}$ is not a one-point measure, $A(r, r'; \mu_{\mathfrak{R}})$ is positive and finite. In fact, both numerator and denominator of (6.19) are positive and finite.

Lemma 6.6 follows from a well-known principal: if R is not a one-point distribution, the covariance with respect to R of strictly increasing functions is positive, (Lehmann 1966).

<u>Lemma 6.7.</u>

Provided $\mu_{\mathfrak{R}}$ is not a one-point measure,

$$(6.20) \qquad \frac{\partial}{\partial r} A(r, r'; \mu_{\mathfrak{R}}) \geq 0$$

and

$$(6.21) \qquad \frac{\partial}{\partial r'} A(r, r'; \mu_{\mathfrak{R}}) \leq 0,$$

with equality if and only if $\mu_{\mathfrak{R}}$ is a two-point measure.

<u>Proof:</u>

The first conclusion of the lemma, (6.20), holds whenever

$$(6.22) \qquad \mathrm{Cov}(\rho, \log(1-\rho)) \frac{\partial}{\partial r} \mathrm{Cov}(\rho, \log \rho) - \mathrm{Cov}(\rho, \log \rho) \frac{\partial}{\partial r} \mathrm{Cov}(\rho, \log(1-\rho)) \geq 0,$$

unless $\mathrm{Cov}(\rho, \log \rho) = 0$, which is excluded since R is not a one-point distribution. After the indicated differentiation, inequality (6.22) becomes:

$$(6.23) \qquad \mathrm{Cov}(\rho,\ \log(1-\rho))E([\rho - E(\rho)][\log \rho - E(\log \rho)]\log \rho)$$
$$- \mathrm{Cov}(\rho,\ \log \rho)E([\rho - E(\rho)][\log(1-\rho) - E(\log(1-\rho))]\log \rho) \geq 0.$$

Letting

$$(6.24) \qquad H(\rho) = [\log \rho - E(\log \rho)]\mathrm{Cov}(\rho,\ \log(1-\rho))$$
$$- [\log(1-\rho) - E(\log(1-\rho))]\mathrm{Cov}(\rho,\ \log \rho),$$

(6.23) can be written,

$$(6.25) \qquad E(H(\rho)[\rho - E(\rho)]\log \rho) \geq 0,$$

or equivalently, since $E(H(\rho)\ [\rho - E(\rho)]) = 0$ according to the definition of H,

$$(6.26) \qquad E(H(\rho)[\rho - E(\rho)][\log \rho - C]) \geq 0,$$

for any constant C.

H is strictly convex since it is the sum of two strictly convex functions, and the expected value of H is zero; therefore, since R is not a one-point distribution, H has exactly two zeros in $(0, 1)$, call them ρ_1 and ρ_2, with $\rho_1 < \rho_2$:

$$(6.27) \qquad H(\rho_1) = H(\rho_2) = 0.$$

Since $H(\rho)$ is convex and $E(H(\rho)) = 0$,

$$(6.28) \qquad \rho_2 > E(\rho) > \rho_1,$$

according to Jensen's inequality, see, for example, (Hardy, et al. 1934, Chapter III).

The number

$$(6.29) \qquad \mathrm{Cov}(H(\rho), \log \rho) = E(H(\rho)[\log \rho - C])$$

$$= \mathrm{Cov}(\log \rho, \log \rho)\mathrm{Cov}(\rho, \log(1-\rho))$$

$$- \mathrm{Cov}(\log \rho, \log(1-\rho))\mathrm{Cov}(\rho, \log \rho),$$

may be of either sign; correspondingly, two cases will be considered.

Case 1:

$\mathrm{Cov}(H(\rho), \log \rho) \geq 0.$

In (6.26), let $C = \log \rho_1$, then

$$(6.30) \qquad E(H(\rho)[\rho - E(\rho)][\log \rho - \log \rho_1]) = E(H(\rho)[\rho - \rho_2][\log \rho - \log \rho_1])$$

$$+ [\rho_2 - E(\rho)][(H(\rho)[\log \rho - \log \rho_1]).$$

The second term of the right-hand side of (6.30) is nonnegative for this case in view of (6.28), and the first term is nonnegative since, according to the following argument,

$$(6.31) \qquad H(\rho)[\rho - \rho_2][\log \rho - \log \rho_1] \geq 0,$$

for all ρ: for $\rho \leq \rho_1$, $H(\rho) \geq 0$, $\rho - \rho_2 < 0$, and $\log \rho - \log \rho_1 \leq 0$; for $\rho_1 < \rho < \rho_2$, $H(\rho) \leq 0$, $\rho - \rho_2 \leq 0$, and $\log \rho - \log \rho_1 \geq 0$; and for $\rho_2 < \rho$, $H(\rho) \geq 0$, $\rho - \rho_2 \geq 0$, and $\log \rho - \log \rho_1 \geq 0$.

Case 2:

$\mathrm{Cov}(H(\rho), \log \rho) < 0.$

In (6.26), let $C = \log \rho_2$, then

$$(6.32)\qquad E(H(\rho)[\rho - E(\rho)][\log \rho - \log \rho_2]) = E(H(\rho)[\rho - \rho_1][\log \rho - \log \rho_2]) + [\rho_1 - E(\rho)]E(H(\rho)[\log \rho - \log \rho_2]).$$

The second term of the right-hand side of (6.32) is nonnegative for this case in view of (6.28), and the first term is nonnegative since, according to the following argument,

$$(6.33)\qquad H(\rho)[\rho - \rho_1][\log \rho - \log \rho_2] \geq 0,$$

for all ρ: for $\rho \leq \rho_1$, $H(\rho) \geq 0$, $\rho - \rho_1 \leq 0$, and $\log \rho - \log \rho_2 < 0$; for $\rho_1 < \rho \leq \rho_2$, $H(\rho) \leq 0$, $\rho - \rho_1 \geq 0$, and $\log \rho - \log \rho_2 \leq 0$; and for $\rho_2 < \rho$, $H(\rho) \geq 0$, $\rho - \rho_1 > 0$, and $\log \rho - \log \rho_2 \geq 0$.

The symmetry of the form of A, given by (6.19), makes it clear that the second conclusion of the lemma, (6.21), is an instance of the first.

In view of (6.15), (6.20) and (6.21) are equalities if $\mu_{\mathfrak{R}}$ is a two-point measure. If $\mu_{\mathfrak{R}}$ is concentrated on more than two points both terms on the right-hand side of (6.30) and (6.32) are positive, so that inequalities (6.20) and (6.21) are strict. □

<u>Proof of Theorem 6.1:</u>

Since, by definition, $\Delta_1(I) = E(\rho) - E(\lambda)$,

$$(6.34)\qquad D_{(a,b)}\Delta_1(r, r'; \mu_{\mathfrak{R}}; L) = D_{(a,b)}E(\rho | r, r'; \mu_{\mathfrak{R}});$$

so that $J(1)$ and $K(1)$ both hold, and Theorem 6.1 determines the sign of the derivative of $\Delta_1(I)$ for every direction in the (r, r') plane from points in the interior of the possibility region for $\mu_{\mathfrak{R}}$.

If $\mu_{\Re}$ is a one-point measure then the theorem is obvious, since a one-point measure is not affected by changes in r or r'; $\mu_{\Re}$ is assumed not to be a one-point measure for the remainder of the proof. This assumption will frequently be used implicitly since Lemmas 6.6 and 6.7 which depend on it will frequently be used.

Equation (4.7) will be used to show $J(n)$ and $K(n)$ inductively. The proof will be acoomplished by considering different combinations of terms in (4.7) that can be simultaneously non-zero. The signs of the last two terms of (4.7) do not materially affect the proof. $\Delta_{n-1}(\sigma R, L)$ can be ≥ 0 or < 0, as can $\Delta_{n-1}(\varphi R, L)$, so that there are four cases to be considered. However, one of these cases is vacuous, as will now be shown.

In view of Lemma 6.6, $\frac{a}{b} > A(r+n-1, r'; \mu_{\Re})$ when $b = 0$ and $\frac{a}{b} < A(r, r'+n-1; \mu_{\Re})$ when $a = 0$, for all n. Therefore, for $n \geq 2$, $J(n-1)$ implies that $\Delta_{n-1}(I)$ does not decrease as r increases and $K(n-1)$ implies that $\Delta_{n-1}(I)$ does not increase as r' increases. In view of these two facts (for $n \geq 2$),

$$(6.35) \qquad \Delta_{n-1}(r+1, r', \mu_{\Re}; L) \geq \Delta_{n-1}(r, r', \mu_{\Re}; L) \geq \Delta_{n-1}(r, r'+1, \mu_{\Re}; L).$$

This relationship implies:

$$(6.36) \qquad \Delta_{n-1}(\sigma R, L) \geq \Delta_{n-1}(\varphi R, L).$$

Inequality (6.36) will be required in a critical point of the proof; for the present it serves to show that the three cases given below are exhaustive:

Case 1:

$\Delta_{n-1}(\sigma R, L) < 0$ (and $\Delta_{n-1}(\varphi R, L) < 0$);

Case 2:

$\Delta_{n-1}(\sigma R, L) \geq 0$ and $\Delta_{n-1}(\varphi R, L) < 0$;

Case 3:

$\Delta_{n-1}(\varphi R, L) \geq 0$ (and $\Delta_{n-1}(\sigma R, L) \geq 0$).

(This proof considers one case at a time for didactic reasons; these three cases could as well be considered simultaneously.)

Case 1:

$\Delta_{n-1}(\sigma R, L) < 0$ and $\Delta_{n-1}(\varphi R, L) < 0$.

According to (4.7),

$$(6.37) \qquad D_{(a,b)}\Delta_n(I) = E(\lambda)D_{(a,b)}\Delta^-_{n-1}(r, r', \mu_{\Re}; \sigma L) + \overline{E}(\lambda)D_{(a,b)}\Delta^-_{n-1}(r, r', \mu_{\Re}; \varphi L),$$

since in the present case neither of the first two terms in (4.7) contributes to the derivative of Δ_n.

If $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\Re})$ then $J(n-1)$ can be applied to show that both terms on the right-hand side of (6.37) are nonnegative. According to Lemma 6.7, $A(r+n-1, r'; \mu_{\Re}) \geq A(r + (n-1) - 1, r'; \mu_{\Re})$; therefore, $\frac{a}{b} \geq A(r + (n-1) - 1, r'; \mu_{\Re})$. Since $D_{(a,b)}\Delta^-_{n-1}(I) \geq 0$ whenever $D_{(a,b)}\Delta_{n-1}(I) \geq 0$, $J(n)$ follows for this case.

Similarly, $K(n)$ follows from $K(n-1)$ for the present case since, according to Lemma 6.7, $\frac{a}{b} \leq A(r, r' + (n-1) - 1; \mu_{\Re})$ whenever $\frac{a}{b} \leq A(r, r' + n - 1; \mu_{\Re})$.

Case 2:

$$\Delta_{n-1}(\sigma R, L) \geq 0 > \Delta_{n-1}(\varphi R, L).$$

In this case, the derivative in (6.37) is adjusted by the addition of two terms which vanish in the former case:

$$(6.38) \qquad E(\rho)D_{(a,b)}\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) + \Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)D_{(a,b)}E(\rho|r, r'; \mu_{\mathfrak{R}}).$$

If $\Delta_{n-1}(\sigma R, L) > 0$ then $\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ in (6.38) can be replaced by $\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$. If $\Delta_{n-1}(\sigma R, L) = 0$, however, $D_{(a,b)}\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ is not necessarily equal to $D_{(a,b)}\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$; it may instead be 0.

If $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}})$ then $J(n-1)$ implies $D_{(a,b)}\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) \geq 0$, since $A(r+n-1, r'; \mu_{\mathfrak{R}}) = A((r+1) + (n-1) - 1, r'; \mu_{\mathfrak{R}})$. Therefore, whenever $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}})$, the first term of (6.38) is nonnegative.

If $\frac{a}{b} \leq A(r, r'+n-1; \mu_{\mathfrak{R}})$ then $\frac{a}{b} \leq A(r+1, r' + (n-1) - 1; \mu_{\mathfrak{R}})$, since according to Lemma 6.7, $A(r, r'+n-1; \mu_{\mathfrak{R}}) \leq A(r, r'+ (n-1) - 1; \mu_{\mathfrak{R}}) \leq A(r+1, r'+ (n-1) - 1; \mu_{\mathfrak{R}})$, and $K(n-1)$ implies $D_{(a,b)}\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) \leq 0$. Therefore, whenever $\frac{a}{b} \leq A(r, r'+n-1; \mu_{\mathfrak{R}})$, the first term of (6.38) is nonpositive.

The second term of (6.38) is nonegative if $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}})$ and nonpositive if $\frac{a}{b} \leq A(r, r'+n-1; \mu_{\mathfrak{R}})$, since $\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) \geq 0$ for the present case and, in view of

$$(6.39) \qquad A(r+n-1, r'; \mu_{\mathfrak{R}}) \geq A(r, r'; \mu_{\mathfrak{R}}) \geq A(r, r'+n-1; \mu_{\mathfrak{R}}),$$

from Lemma 6.7, J(1) and K(1) imply that $D_{(a,b)}E(\rho|r, r'; \mu_{\mathfrak{R}})$ has the appropriate sign.

Case 3:

$\Delta_{n-1}(\varphi R, L) \geq 0$ and $\Delta_{n-1}(\sigma R, L) \geq 0$.

In this case, the derivative in (6.37) is adjusted by the addition of three terms:

$$(6.40)\qquad E(\rho)D_{(a,b)}\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) + \overline{E}(\rho)D_{(a,b)}\Delta^{+}_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)$$
$$+ [\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) - \Delta^{+}_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)]$$
$$\cdot D_{(a,b)}E(\rho|r, r'; \mu_{\mathfrak{R}}).$$

If $\Delta_{n-1}(\varphi R, L) > 0$ then $\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ and $\Delta^{+}_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)$ in (6.40) can be replaced with $\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ and $\Delta_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)$. If $\Delta_{n-1}(\varphi R, L) = 0$, however, $D_{(a,b)}\Delta^{+}_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ and $D_{(a,b)}\Delta^{+}_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)$ may be equal to $D_{(a,b)}\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)$ and $D_{(a,b)}\Delta_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)$ or either may be 0.

The first term of (6.40) is the same as the first term of (6.38) so that it has the appropriate sign according to the argument given in the previous case.

If $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}})$ then $\frac{a}{b} \geq A(r + (n-1) - 1, r'+1; \mu_{\mathfrak{R}})$, since according to Lemma 6.7, $A(r+n-1, r'; \mu_{\mathfrak{R}}) \geq A(r + (n-1) - 1, r'; \mu_{\mathfrak{R}}) \geq A(r + (n-1) - 1, r'+1; \mu_{\mathfrak{R}})$, and J(n-1) implies $D_{(a,b)}\Delta_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L) \geq 0$. Therefore, whenever $\frac{a}{b} \geq A(r+n-1, r'; \mu_{\mathfrak{R}})$, the second term of (6.40) is nonnegative.

If $\frac{a}{b} \geq A(r, r'+n-1; \mu_{\Re})$ then $K(n-1)$ implies $D_{(a,b)}\Delta_{n-1}(r, r'+1, \mu_{\Re}; L) \leq 0$, since $A(r, r'+n-1; \mu_{\Re}) = A(r, (r'+1) + (n-1) - 1; \mu_{\Re})$. Therefore, whenever $\frac{a}{b} \leq A(r, r'+n-1; \mu_{\Re})$, the second term of (6.40) is nonpositive.

In this case, the third term of (6.40) is nonnegative if $\frac{a}{b} \leq A(r, r'+n-1; \mu_{\Re})$, since $\Delta^+_{n-1}(r+1, r', \mu_{\Re}; L) - \Delta^+_{n-1}(r, r'+1, \mu_{\Re}; L) \geq 0$ according to (6.18) and, in view of (6.39), $J(1)$ and $K(1)$ imply that $D_{(a,b)}E(\rho|r, r'; \mu_{\Re})$ has the appropriate sign.

$J(n)$ and $K(n)$ follow from $J(n-1)$ and $K(n-1)$ for all three cases and the theorem is proved. $\square$

7. Fundamental inequalities; extensions

Theorem 6.1, which deals with the sign of the gradient of $\Delta_n(r,r',\mu_{\Re};\, L)$ along curves in the interior of the possibility region for $\mu_{\Re}$, is extended by the three theorems of the present section. Theorem 7.1 is a macroscopic version of Theorem 6.1, in which the contours of $E(\rho\,|\,r+n-1,\, r';\, \mu_{\Re})$ and $E(\rho\,|\,r,r'+n-1;\, \mu_{\Re})$ are shown to be lines of nondecrease and nonincrease of $\Delta_n(r,r',\mu_{\Re};\, L)$ in any direction of nondecreasing r and r' in the interior of the possibility region for $\mu_{\Re}$. Theorem 7.2 extends Theorem 7.1 to include the edges of the possibility region for $\mu_{\Re}$; the region may have no edges, one edge, or two edges. Finally, Theorem 7.3 shows that for fixed $\mu_{\Re}$, L, and n, $\Delta_n(r,r',\mu_{\Re};\, L)$ is strictly increased or decreased if $E(\rho\,|\,r+n-1,\, r';\, \mu_{\Re})$ or $E(\rho\,|\,r,r'+n-1;\, \mu_{\Re})$ is increased or decreased.

Theorem 7.1.

Provided

$$(7.1)\qquad \nu(r+\delta r,\, r'+\delta r';\, \mu_{\Re}) < \infty \quad \text{for some} \quad \delta r,\, \delta r' < 0,$$

the following statements are true for $n \geq 1$, for $I = (r,r',\mu_{\Re};\, L)$, and for all $\delta r,\, \delta r' \geq 0$:

$\hat{J}(n)$: $\Delta_n(r+\delta r,\, r'+\delta r',\, \mu_{\Re};\, L) \geq \Delta_n(r,r',\mu_{\Re};\, L)$

$$\text{if} \quad E(\rho\,|\,r+\delta r+n-1,\, r'+\delta r';\, \mu_{\Re}) \geq E(\rho\,|\,r+n-1,\, r';\, \mu_{\Re});$$

$\hat{K}(n)$: $\Delta_n(r+\delta r,\, r'+\delta r',\, \mu_{\Re};\, L) \leq \Delta_n(r,r',\mu_{\Re};\, L)$

$$\text{if} \quad E(\rho\,|\,r+\delta r,\, r'+\delta r'+n-1;\, \mu_{\Re}) \leq E(\rho\,|\,r,\, r'+n-1;\, \mu_{\Re}).$$

Proof:

The implicit function theorem (Widder 1961, Theorem 14, p. 56) applies to show that on expressing the contours of $E(\rho|r, r'; \mu_{\mathfrak{R}})$ as $(r(N_{\mathfrak{R}}), r'(N_{\mathfrak{R}}))$ in the parameter $N_{\mathfrak{R}} = r + r'$, each contour extends uninterrupted for all $N_{\mathfrak{R}}$; the slope of the contours of $E(\rho|r, r'; \mu_{\mathfrak{R}})$ is $A(r, r'; \mu_{\mathfrak{R}})$, which is defined by (6.13).

For $\delta r' \geq 0$, consider two points $(r, r'; \mu_{\mathfrak{R}})$ and $(r+h, r'+\delta r'; \mu_{\mathfrak{R}})$ on a contour of $E(\rho|x+n-1, x'; \mu_{\mathfrak{R}})$; that is,

$$(7.2) \qquad E(\rho|r+n-1, r'; \mu_{\mathfrak{R}}) = E(\rho|r+h+n-1, r'+\delta r'; \mu_{\mathfrak{R}}).$$

According to $J(n)$ of Theorem 6.1, $\Delta_n(x, x', \mu_{\mathfrak{R}}; L)$ is nondecreasing along such a contour in the interior of the possibility region for $\mu_{\mathfrak{R}}$ for any L, so that

$$(7.3) \qquad \Delta_n(r+h, r'+\delta r', \mu_{\mathfrak{R}}; L) \geq \Delta_n(r, r', \mu_{\mathfrak{R}}; L).$$

Consider a third point $(r+\delta r, r'+\delta r'; \mu_{\mathfrak{R}})$ satisfying the condition in $\hat{J}(n)$, so that

$$(7.4) \qquad E(\rho|r+\delta r+n-1, r'+\delta r'; \mu_{\mathfrak{R}}) \geq E(\rho|r+h+n-1, r'+\delta r'; \mu_{\mathfrak{R}})$$

according to (7.2). According to $J(n)$ of Theorem 6.1 for $b = 0$,

$$(7.5) \qquad \Delta_n(r+\delta r, r'+\delta r', \mu_{\mathfrak{R}}; L) \geq \Delta_n(r+h, r'+\delta r', \mu_{\mathfrak{R}}; L).$$

$\hat{J}(n)$ of the theorem follows from (7.3) and (7.5).

$\hat{K}(n)$ of the theorem is proved in a similar fashion by considering points on a contour of $E(\rho|x, x'+n-1; \mu_{\mathfrak{R}})$ and applying $K(n)$ of Theorem 6.1. □

Theorem 7.1 does not apply for any distribution $R = (r,r'; \mu_{\Re})$ which corresponds to a point on an edge of the possibility region for $\mu_{\Re}$; for such a distribution, $\nu(r+\delta r, r'+\delta r'; \mu_{\Re}) = \infty$ if $\delta r < 0$ or if $\delta r' < 0$ depending on whether (r, r') is on the vertical edge $(r = r_*)$ or horizontal edge $(r' = r_*')$. Theorem 7.1 will be extended to arbitrary distributions $R = (r,r'; \mu_{\Re})$ by showing first that $\Delta_n(R, L)$ can be approximated arbitrarily closely by replacing $\mu_{\Re}$ with a measure which satisfies condition (7.1) in Theorem 7.1.

<u>Lemma 7.1.</u>

For each $I = (r_0, r_0', \mu_{\Re}; L)$ for which (r_0, r_0') is in the possibility region of $\mu_{\Re}$ and for which $\mu_{\Re}$ is not confined to the two points $\{0, 1\}$, there exists a family of measures m_ε such that for all real r and r',

$$(7.6) \qquad \nu(r, r'; m_\varepsilon) < \infty$$

for each measure m_ε with $\varepsilon > 0$, and

$$(7.7) \qquad \lim_{\varepsilon \downarrow 0} \Delta_n(r,r',m_\varepsilon; L) = \Delta_n(r,r',\mu_{\Re}; L)$$

for $n \geq 1$ and every r and r' for which $r \geq r_0$ and $r' \geq r_0'$.

<u>Remarks.</u>

The convergence in (7.7) is not necessarily uniform in (r, r') or in n.

Any measure which satisfies (7.6) also satisfies (7.1), so that for $\varepsilon > 0$ Theorem 7.1 applies to $I = (r,r',m_\varepsilon; L)$.

Proof of Lemma 7.1:

To prove the lemma a family of measures m_ε which depends on $\mu_{\mathfrak{R}}$ and for which each of the m_ε with $\varepsilon > 0$ satisfies (7.6) has to be exhibited. For $\varepsilon < \frac{1}{2}$, m_ε will be constructed from $\mu_{\mathfrak{R}}$ in a completely explicit and very simple way; the definition of m_ε for $\varepsilon \geq \frac{1}{2}$ is of course almost immaterial, so for $\varepsilon \geq \frac{1}{2}$ let m_ε concentrate measure 1 on the point $\frac{1}{2}$. It is enough to prove the lemma for $r_0 = r_0' = 0$, because if $d\mu_{\mathfrak{R}}^*(\rho) = \rho^{r_0}(1-\rho)^{r_0'} d\mu_{\mathfrak{R}}(\rho)$, then $(r, r', \mu_{\mathfrak{R}}; L) = (r-r_0, r'-r_0', \mu_{\mathfrak{R}}^*; L)$.

In fact, for $r_0 = r_0' = 0$, m_ε (for $\varepsilon \leq \frac{1}{2}$) will be the result of shrinking $\mu_{\mathfrak{R}}$ toward $\rho = \frac{1}{2}$ by the factor $1 - 2\varepsilon$. For any set $S \subset [0, 1]$ and $0 < \varepsilon < \frac{1}{2}$ define the set S_ε to be $\varepsilon + (1-2\varepsilon)S$ in the usual algebraic sense, so that for all $\rho \in [0, 1]$,

$$(7.8) \qquad \rho \in S_\varepsilon \text{ iff } \frac{\rho - \varepsilon}{1 - 2\varepsilon} \in S.$$

Define

$$(7.9) \qquad m_\varepsilon(S) = \mu_{\mathfrak{R}}(S_\varepsilon),$$

for any set $S \subset [0, 1]$ such that S is Borel measurable. Then

$$(7.10) \qquad m_\varepsilon([0, \varepsilon)) = m_\varepsilon((1-\varepsilon, 1]) = 0,$$

and (7.6) holds as long as $\varepsilon > 0$.

To see that (7.7) holds at $n = 1$ for the family of measures defined by (7.9), write

$$(7.11)\qquad \nu(r,r';\, m_\varepsilon) = \int_0^1 \rho^r(1-\rho)^{r'} dm_\varepsilon(\rho)$$

$$= \int_\varepsilon^{1-\varepsilon} \rho^r(1-\rho)^{r'} d\mu_{\Re}\left(\frac{\rho-\varepsilon}{1-2\varepsilon}\right)$$

$$= \int_0^1 [x + \varepsilon(1-2x)]^r[1 - x - \varepsilon(1-2x)]^{r'} d\mu_{\Re}(x)$$

$$= \int_0^1 x^r(1-x)^{r'} d\mu_{\Re}(x) + o(1)$$

for $r, r' \geq 0$ according to the Lebesgue dominated convergence theorem. Therefore,

$$(7.12)\qquad E(\rho|r,r';\, m_\varepsilon) = \frac{\nu(r+1,\, r';\, m_\varepsilon)}{\nu(r,\, r';\, m_\varepsilon)}$$

$$= \frac{\nu(r+1,\, r';\, \mu_{\Re}) + o(1)}{\nu(r,\, r';\, \mu_{\Re}) + o(1)}$$

$$= E(\rho|r,\, r';\, \mu_{\Re}) + o(1),$$

which proves (7.7) for $n = 1$.

For $n \geq 2$, in view of (4.7) and (7.11),

$$\begin{aligned}(7.13)\qquad \Delta_n(r,r',m_\varepsilon;\, L) &= E(\rho|r,r';\, m_\varepsilon)\Delta^+_{n-1}(r+1,\, r',\, m_\varepsilon;\, L) \\ &\quad + \bar{E}(\rho|r,r';\, m_\varepsilon)\Delta^+_{n-1}(r,\, r'+1,\, m_\varepsilon;\, L) \\ &\quad + E(\lambda)\Delta^-_{n-1}(r,r',m_\varepsilon;\, \sigma L) + \bar{E}(\lambda)\Delta^-_{n-1}(r,r',m_\varepsilon;\varphi L) \\ &= E(\rho|r,r';\, \mu_{\Re})\Delta^+_{n-1}(r+1,\, r',\, m_\varepsilon;\, L) \\ &\quad + \bar{E}(\rho|r,r';\, \mu_{\Re})\Delta^+_{n-1}(r,\, r'+1,\, m_\varepsilon;\, L) \\ &\quad + E(\lambda)\Delta^-_{n-1}(r,r',m_\varepsilon;\, \sigma L) \\ &\quad + \bar{E}(\lambda)\Delta^-_{n-1}(r,r',m_\varepsilon;\, \varphi L) + o(1).\end{aligned}$$

Assuming that (7.7) holds at $n-1$ for $r, r' \geq 0$ and for all L and, in view of (7.13),

$$\begin{aligned}(7.14)\qquad \lim_{\varepsilon\downarrow 0} \Delta_n(r, r', m_\varepsilon; L) &= E(\rho|r,r'; \mu_{\mathfrak{R}})\Delta^+_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L)\\ &\quad + \overline{E}(\rho|r,r'; \mu_{\mathfrak{R}})\Delta^+_{n-1}(r, r'+1, \mu_{\mathfrak{R}}; L)\\ &\quad + E(\lambda)\Delta^-_{n-1}(r, r', \mu_{\mathfrak{R}}; \sigma L)\\ &\quad + \overline{E}(\lambda)\Delta^-_{n-1}(r, r', \mu_{\mathfrak{R}}; \varphi L)\\ &= \Delta_n(r, r', \mu_{\mathfrak{R}}; L),\end{aligned}$$

therefore (7.7) holds at n. □

Lemma 7.1 will be used to prove the next theorem, which extends Theorem 7.1 to include arbitrary measures. The proof of $\hat{J}(n)$ depends only on $\hat{J}(n)$ of Theorem 7.1 and the proof of $\hat{K}(n)$ depends only on $\hat{K}(n)$ of Theorem 7.1.

Theorem 7.2.

The following statements are true for $n \geq 1$, for all $I = (r,r',\mu_{\mathfrak{R}}; L)$, and for δr, $\delta r' \geq 0$:

$\hat{J}(n)$: $\Delta_n(r+\delta r, r'+\delta r', \mu_{\mathfrak{R}}; L) \geq \Delta_n(r,r',\mu_{\mathfrak{R}}; L)$

if $E(\rho|r+\delta r+n-1, r'+\delta r'; \mu_{\mathfrak{R}}) \geq E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$;

$\hat{K}(n)$: $\Delta_n(r+\delta r, r'+\delta r', \mu_{\mathfrak{R}}; L) \leq \Delta_n(r,r',\mu_{\mathfrak{R}}; L)$

if $E(\rho|r+\delta r, r'+\delta r'+n-1; \mu_{\mathfrak{R}}) \leq E(\rho|r, r'+n-1; \mu_{\mathfrak{R}})$.

Remark.

It was noted in Section 2 that for all δr, $\delta r' \geq 0$, $(r+\delta r, r'+\delta r')$ is in the possibility region for $\mu_{\mathfrak{R}}$ whenever (r, r') is, unless $\mu_{\mathfrak{R}}((0, 1)) = 0$, that is, unless $\mu_{\mathfrak{R}}(0) + \mu_{\mathfrak{R}}(1) = 1$. In the latter event the possibility region for $\mu_{\Re}$ consists at most of the nonnegative axes. For such measures $\hat{J}(n)$ and $\hat{K}(n)$ of Theorem 7.2 may be meaningless, depending on δr and $\delta r'$. The convention is adopted here that $\hat{J}(n)$ and $\hat{K}(n)$ have content only if $(r+\delta r,\ r'+\delta r')$ is in the possibility region for $\mu_{\Re}$. This convention does not exclude the extreme one-point or two-point measures from consideration in the theorem, but it does eliminate consideration of any direction out from the possibility region for $\mu_{\Re}$. These easy special cases are not explicitly covered in the proof below.

<u>Proof of Theorem 7.2:</u>

Assume that R is a distribution for which $\hat{J}(n)$ is false; say for $\delta r,\ \delta r' \geq 0$,

$$E(\rho\,|\,r+\delta r+n-1,\ r'+\delta r';\ \mu_{\Re}) - E(\rho\,|\,r+n-1,\ r';\ \mu_{\Re}) \geq 0, \tag{7.15}$$

while

$$\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ L) - \Delta_n(r, r', \mu_{\Re};\ L) < 0. \tag{7.16}$$

Unless $\mu_{\Re}$ is a one-point measure, in which case the theorem is already known to hold, or unless $\mu_{\Re}((0,\ 1)) = 0$, which is a case not currently under discussion, if (7.15) and (7.16) can hold at all, they hold with strict inequality in (7.15), as will now be argued. Either δr or $\delta r'$ is positive. Say for definiteness that $\delta r > 0$; the other possibility is very similar. If δr is replaced by a slightly larger value (7.16) will not be lost; for $\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ L)$ is continuous in δr for $\delta r > 0$, according to a slight variant of Lemma 6.2. But if δr is increased, (7.15) will be rendered strict according to Lemma 6.7 (the proof of which requires no greater generality than is at hand).

In view of (7.16) and Lemma 7.1 there is a measure m_ε which satisfies (7.1) and which approximates $\mu_{\mathfrak{R}}$ sufficiently well to guarantee that

$$(7.17) \qquad \Delta_n(r+\delta r,\ r'+\delta r',\ m_\varepsilon;\ L) - \Delta_n(r,r',m_\varepsilon;\ L) < 0$$

for sufficiently small ε and also, since (7.15) is now supposed to hold with strict inequality,

$$(7.18) \qquad E(\rho\,|\,r+\delta r+n-1,\ r'+\delta r';\ m_\varepsilon) - E(\rho\,|\,r+n-1,\ r';\ m_\varepsilon) > 0.$$

This contradicts $\hat{J}(n)$ of Theorem 7.1.

A similar argument delivers $\hat{K}(n)$. □

The next theorem strengthens Theorem 7.2 to show that a strict increase in $E(\rho\,|\,r+n-1,\ r';\ \mu_{\mathfrak{R}})$ or a strict decrease in $E(\rho\,|\,r,\ r'+n-1;\ \mu_{\mathfrak{R}})$ guarantees a strict increase or decrease in $\Delta_n(r,\ r',\ \mu_{\mathfrak{R}};\ L)$ for all L and n.

<u>Theorem 7.3</u>.

The following statements are true for $n \geq 1$, for all $I = (r,\ r',\ \mu_{\mathfrak{R}};\ L)$, and for $\delta r,\ \delta r' \geq 0$:

$J^*(n)$: $\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\mathfrak{R}};\ L) > \Delta_n(r,\ r',\ \mu_{\mathfrak{R}};\ L)$

if $E(\rho\,|\,r+\delta r+n-1,\ r'+\delta r';\ \mu_{\mathfrak{R}}) > E(\rho\,|\,r+n-1,\ r';\ \mu_{\mathfrak{R}})$;

$K^*(n)$: $\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\mathfrak{R}};\ L) < \Delta_n(r,\ r',\ \mu_{\mathfrak{R}};\ L)$

if $E(\rho\,|\,r+\delta r,\ r'+\delta r'+n-1;\ \mu_{\mathfrak{R}}) < E(\rho\,|\,r,\ r'+n-1;\ \mu_{\mathfrak{R}})$.

<u>Remarks</u>.

The theorem is true for all distributions $R = (r,\ r';\ \mu_{\mathfrak{R}})$, but the conditions in $J^*(n)$ and $K^*(n)$ clearly indicate that the theorem has no content if $\mu_{\mathfrak{R}}$ is a one-point measure, for in that case R is not affected by changes is r or r'.

The proof of Theorem 7.3 can be viewed as a modification of the proof of Theorem 6.1 (with differences playing the role of derivatives). The key to the modification is the demonstration that under the condition in $J^*(n)$ or in $K^*(n)$ the four terms in (4.7) cannot vanish simultaneously--they may all vanish if R is a one-point distribution (but only when L is the same one-point distribution).

Like Theorem 7.2, Theorem 7.3 can easily be interpreted as true in case $\mu_{\mathfrak{R}}$ is confined to the two extreme points $\{0, 1\}$, but this possibility will not be attended to in the following proof.

Suppose δr and $\delta r'$ are positive and equality holds in the second comparison in $J^*(n)$ (or in $K^*(n)$), can the first inequality nonetheless be concluded? No, not if $\mu_{\mathfrak{R}}$ is carried by at most two points, as the attentive reader may perceive (in view of Lemma 6.7), but otherwise it does, though this extension of the theorem will not be carried out in the present paper.

<u>Proof of Theorem 7.3</u>:

The theorem will be proved by induction, starting at $n = 1$, where it is trivial.

In view of (4.7), for $n \geq 2$,

$$(7.19)\qquad \begin{aligned} &\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ L) - \Delta_n(r,\ r',\ \mu_{\Re};\ L) \\ &\quad = E(\rho\,|\,r+\delta r,\ r'+\delta r';\ \mu_{\Re})\Delta^{+}_{n-1}(r+\delta r+1,\ r'+\delta r',\ \mu_{\Re};\ L) \\ &\qquad + \overline{E}(\rho\,|\,r+\delta r,\ r'+\delta r';\ \mu_{\Re})\Delta^{+}_{n-1}(r+\delta r,\ r'+\delta r'+1,\ \mu_{\Re};\ L) \\ &\qquad + E(\lambda)\Delta^{-}_{n-1}(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ \sigma L) \\ &\qquad + \overline{E}(\lambda)\Delta^{-}_{n-1}(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ \varphi L) \\ &\qquad - E(\rho\,|\,r,\ r';\ \mu_{\Re})\Delta^{+}_{n-1}(r+1,\ r',\ \mu_{\Re};\ L) \\ &\qquad - \overline{E}(\rho\,|\,r,\ r';\ \mu_{\Re})\Delta^{+}_{n-1}(r,\ r'+1,\ \mu_{\Re};\ L) \\ &\qquad - E(\lambda)\Delta^{-}_{n-1}(r,\ r',\ \mu_{\Re};\ \sigma L) \\ &\qquad - \overline{E}(\lambda)\Delta^{-}_{n-1}(r,\ r',\ \mu_{\Re};\ \varphi L). \end{aligned}$$

The right side of (7.19) can be rewritten:

$$(7.20)\qquad \begin{aligned} &E(\rho\,|\,r,\ r';\ \mu_{\Re})[\Delta^{+}_{n-1}(r+\delta r+1,\ r'+\delta r',\ \mu_{\Re};\ L) - \Delta^{+}_{n-1}(r+1,\ r',\ \mu_{\Re};\ L)] \\ &\quad + \overline{E}(\rho\,|\,r,\ r';\ \mu_{\Re})[\Delta^{+}_{n-1}(r+\delta r,\ r'+\delta r'+1,\ \mu_{\Re};\ L) - \Delta^{+}_{n-1}(r,\ r'+1,\ \mu_{\Re};\ L)] \\ &\quad + (E(\rho\,|\,r+\delta r,\ r'+\delta r';\ \mu_{\Re}) - E(\rho\,|\,r,\ r';\ \mu_{\Re})) \\ &\qquad \cdot\ [\Delta^{+}_{n-1}(r+\delta r+1,\ r'+\delta r',\ \mu_{\Re};\ L) - \Delta^{+}_{n-1}(r+\delta r,\ r'+\delta r'+1,\ \mu_{\Re};\ L)] \\ &\quad + E(\lambda)[\Delta^{-}_{n-1}(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ \sigma L) - \Delta^{-}_{n-1}(r,\ r',\ \mu_{\Re};\ \sigma L)] \\ &\quad + \overline{E}(\lambda)[\Delta^{-}_{n-1}(r+\delta r,\ r'+\delta r',\ \mu_{\Re};\ \varphi L) - \Delta^{-}_{n-1}(r,\ r',\ \mu_{\Re};\ \varphi L)]. \end{aligned}$$

For $n \geq 2$, $J^{*}(n-1)$ and $K^{*}(n-1)$ apply to show that

$$(7.21)\qquad \Delta_{n-1}(\sigma R,\ L) > \Delta_{n-1}(R,\ L) > \Delta_{n-1}(\varphi R,\ L),$$

which is a strict inequality version of (6.35). (7.21) implies that at least one term of the right side of (4.7) is nonzero (cf. Theorem 5.4).

For, in view of (7.21), Theorem 7.2, and the fact that $\Delta_n(R, L) = -\Delta_n(L, R)$,

$$(7.22) \qquad \Delta_{n-1}(\sigma R, L) > \Delta_{n-1}(R, L) = -\Delta_{n-1}(L, R) \geq$$
$$\geq \Delta_{n-1}(\sigma L, R) = \Delta_{n-1}(R, \sigma L);$$

therefore, either $\Delta_{n-1}(\sigma R, L) > 0$ or $\Delta_{n-1}(R, \sigma L) < 0$.

Assume $E(\rho|r+\delta r+n-1, r'+\delta r'; \mu_{\mathfrak{R}}) \geq E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$, then $J^*(n-1)$ implies that the bracketed portion of the first term of (7.20) is positive when $\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) > 0$ and the bracketed portion of the fourth term of (7.20) is positive when $\Delta_{n-1}(r, r', \mu_{\mathfrak{R}}; \sigma L) < 0$. Therefore, the first term of (7.20) is positive when $\Delta_{n-1}(r+1, r', \mu_{\mathfrak{R}}; L) > 0$ since $E(\rho|r, r'; \mu_{\mathfrak{R}})$ cannot then be zero and the fourth term of (7.20) is positive when $\Delta_{n-1}(r, r', \mu_{\mathfrak{R}}; \sigma L) < 0$ since $E(\lambda)$ cannot then be zero. In either case the remaining terms of (7.20) are nonnegative in view of $\hat{J}(n-1)$ of Theorem 7.2, so that $J^*(n-1)$ implies $J^*(n)$.

A similar argument uses $K^*(n-1)$ of the theorem and $\hat{K}(n-1)$ of Theorem 7.2 to deliver $K^*(n)$. □

8. Results that hold for all n

In this section, the inequalities derived in the previous section will be used to examine parts of the domain space where the sign $\Delta_n(I)$ is the same for all n--Theorem 8.3 is the only result in this section which depends on n. The conclusions rest on the principal theorems, Theorems 7.2 and 7.3, but only for the special cases $\delta r' = 0$ in $\hat{J}(n)$ and $J^*(n)$ and $\delta r = 0$ in $\hat{K}(n)$ and $K^*(n)$. These theorems will be used in their full generality in the next section.

Theorem 8.1.

For all $I = (R, L)$ and $n \geq 2$,

$$(8.1) \qquad \Delta_n(R, L) \leq \Delta^+_{n-1}(\sigma R, L),$$

with strict inequality if $0 \leq \Delta_n(R, L)$ and R is not a one-point distribution.

Proof:

According to (4.7),

$$(8.2) \qquad \Delta_n(R, L) \leq E(\rho)\Delta^+_{n-1}(\sigma R, L) + \overline{E}(\rho)\Delta^+_{n-1}(\varphi R, L),$$

and the inequality is strict unless $0 \leq \Delta_{n-1}(R, \sigma L)$. The right side of (8.2) is

$$(8.3) \qquad \Delta^+_{n-1}(\sigma R, L) - \overline{E}(\rho)(\Delta^+_{n-1}(\sigma R, L) - \Delta^+_{n-1}(\varphi R, L)) \leq \Delta^+_{n-1}(\sigma R, L),$$

in view of (6.36), which holds for all R according to Theorem 7.2. In view of (7.21), the strict-inequality version of (6.36), inequality (8.3) is strict when R is not a one-point distribution unless $\Delta_{n-1}(\sigma R, L) \leq 0$. But if R is not a one-point distribution,

$\Delta_{n-1}(\sigma R, L) \leq 0 \leq \Delta_{n-1}(R, \sigma L)$ cannot be satisfied in view of (7.26); therefore, in this case either inequality (8.2) or inequality (8.3) is strict. □

Bradt, et al. (1956) prove the following result (which they call the "stay-on-a-winner-rule") for the one-armed bandit problem. (A two-armed bandit is called a <u>one-armed bandit</u> if either ρ or λ is known with probability one; that is, if R or L is a one-point distribution.) Quisel (1965) offers a proof of this result for the two-armed bandit that is different from the present proof. Theorem 8.2 is a corollary of Theorem 8.1; it means that if an arm is optimal and pulled and yields a success, then it is optimal on the next pull as well.

<u>Theorem 8.2</u>.

For all $I = (R, L)$ for which R is not a one-point distribution and $n \geq 2$, $\Delta_n(R, L) \geq 0$ implies $\Delta_{n-1}(\sigma R, L) > 0$; if R is a one-point distribution then $\Delta_n(R, L) \geq 0$ implies $\Delta_{n-1}(\sigma R, L) \geq 0$ and $\Delta_n(R, L) > 0$ implies $\Delta_{n-1}(\sigma R, L) > 0$.

<u>Proof</u>:

Immediate from Theorem 8.1. □

Nothing can be said in general about the relationship between $\Delta_n(R, L)$ and $\Delta_{n-1}(\varphi R, L)$; either can be less than the other. For example, suppose $n = 2$ and L is determined by $\mu_{\mathfrak{L}}(\lambda) = \beta(\lambda) = \lambda^{-1}(1-\lambda)^{-1}$ and $\ell = \ell' = 1$. If R is such that $\mu_{\mathfrak{R}} = \beta$ and $r = r' = 1/2$, then using the notation $N_{\mathfrak{R}} = r + r'$ and $N_{\mathfrak{L}} = \ell + \ell'$, according to (5.10), case (+0-0),

$$(8.4) \qquad \Delta_2(R, L) = E(\rho^2) - E(\lambda^2) = \frac{r+1}{N_{\mathfrak{R}}+1} - \frac{\ell+1}{N_{\mathfrak{L}}+1} = \frac{3}{4} - \frac{2}{3} = \frac{1}{12} ;$$

while, according to (4.8),

$$(8.5) \qquad \Delta_1(\varphi R, L) = \frac{E(\rho) - E(\rho^2)}{1 - E(\rho)} - E(\lambda) = \frac{r}{N_{\mathcal{R}}+1} - \frac{\ell}{N_{\mathcal{L}}}$$

$$= \frac{1}{4} - \frac{1}{2} = -\frac{1}{4} < \Delta_2(R, L).$$

If however, $r = 11$ and $r' = 9$, then according to (5.10), case (++-0),

$$(8.6) \qquad \Delta_2(R, L) = E(\rho) - E(\lambda) - E(\lambda^2) + E(\rho)E(\lambda)$$

$$= \frac{r}{N_{\mathcal{R}}} - \frac{\ell}{N_{\mathcal{L}}} - \frac{\ell}{N_{\mathcal{L}}}\frac{\ell+1}{N_{\mathcal{L}}+1} + \frac{r}{N_{\mathcal{R}}}\frac{\ell}{N_{\mathcal{L}}}$$

$$= \frac{11}{20} - \frac{1}{2} - \frac{1}{2}\frac{2}{3} + \frac{11}{20}\frac{1}{2} = -\frac{1}{120};$$

while,

$$(8.7) \qquad \Delta_1(\varphi R, L) = \frac{r}{N_{\mathcal{R}}+1} - \frac{\ell}{N_{\mathcal{L}}} = \frac{11}{21} - \frac{1}{2} = \frac{1}{42} > \Delta_2(R, L).$$

That $\Delta_n(R, L)$ and $\Delta_{n-1}(\varphi R, L)$ are not related in the way that $\Delta_n(R, L)$ and $\Delta_{n-1}(\sigma R, L)$ are is a manifestation of the asymmetry of the two-armed bandit problem in successes and failures, an asymmetry not evinced by Theorems 7.2 and 7.3. Heuristically, a success on an optimal arm never decreases (and typically increases) the inclination to pull that arm again, while a failure on an optimal arm (obviously could decrease, but also) can increase the inclination to pull the arm again. This is because the number of pulls remaining has been lessened by 1 leaving less time to take advantage of anything learned.

The one-armed bandit can be instructive in this regard. Suppose that R is a one-point distribution and that L is not a one-point distribution. In this case $\varphi R = \sigma R = R$ and (8.1) yields

$$\Delta_n(R, L) \leq \Delta^+_{n-1}(\varphi R, L) = \Delta^+_{n-1}(R, L). \tag{8.8}$$

It can easily happen that the left arm is worth pulling on the first of n pulls remaining on the chance that it is really better than the right arm, and not worth pulling on the first of $n - 1$ pulls remaining. The latter example above (with calculations in (8.6) and (8.7)) is much like a one-armed bandit since $N_{\mathfrak{R}}$ is large relative to $N_{\mathfrak{L}}$.

The next theorem, Theorem 8.3, is the only result in this section which depends on n, but it is really a corollary of Theorem 8.1 which is true for all n. The intuitive notion of Theorem 8.3 is that an arm should be pulled at the last stage (that is, when $n = 1$) if it was optimal at some previous stage and has since yielded all successes. The theorem gives a crude but easily computable sufficient condition on the distributions R and L for the optimality of $\mathfrak{L}$; and, of course, there is a symmetric condition for the optimality of $\mathfrak{R}$.

<u>Theorem 8.3</u>.

For all n and $I = (r, r', \mu_{\mathfrak{R}}; L)$, if

$$E(\lambda) \geq E(\rho \mid r+n-1, r'; \mu_{\mathfrak{R}}), \tag{8.9}$$

then $\Delta_n(I) \leq 0$, and $\Delta_n(I) < 0$ if R is not a one-point distribution.

<u>Remark</u>.

If $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \beta$, then condition (8.9) becomes:

$$\frac{l}{N_{\mathfrak{L}}} \geq \frac{r + n - 1}{N_{\mathfrak{R}} + n - 1}\ . \tag{8.10}$$

Proof of Theorem 8.3:

Assuming (8.9) and applying Theorem 8.1 $n - 1$ times,

$$(8.11) \qquad 0 \geq E(\rho|r+n-1, r'; \mu_{\mathfrak{R}}) - E(\lambda) = \Delta_1(\sigma^{n-1}R, L) \\ \geq \Delta_2^+(\sigma^{n-2}R, L) \geq \ldots \geq \Delta_n(R, L),$$

with strict inequality if R is not a one-point distribution. □

Condition (8.9) is more easily satisfied for small n since $E(\rho|r+n-1, r'; \mu_{\mathfrak{R}})$ is nondecreasing (and typically increasing) in n. Moreover, if R associates positive probability to all intervals $(1-\epsilon, 1]$, $\epsilon > 0$, then

$$(8.12) \qquad \lim_{n\to\infty} E(\rho|r+n-1, r'; \mu_{\mathfrak{R}}) = 1,$$

and if R is such a distribution, (8.9) would be satisfied for very large n only if, under L, $\lambda = 1$ with probability one. For fixed n and $E(\lambda)$, (8.9) is more easily satisfied for distributions R that concentrate probability near $E(\rho)$; for example, if R is a one-point distribution, $E(\rho|r+n-1, r'; \mu_{\mathfrak{R}}) = E(\rho|r,r'; \mu_{\mathfrak{R}})$ and the problem is a one-armed bandit, then Theorem 8.3 impliestthat $\mathfrak{L}$ is optimal for all n whenever $E(\lambda) \geq E(\rho)$. This application of Theorem 8.3 is intuitive since a left arm which will yield at least as much immediate expected income and at least as much information as the right arm would seem to be the optimal arm.

In the remaining results of this section, R and L are assumed to be conjugate with respect to each other; that is, given R and L there exist $\mu_{\mathfrak{R}}$ and $\mu_{\mathfrak{L}}$ such that $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}}$. The following result means that whenever one of the two comparable arms has a greater "effective number" of successes (given by r and l) and a smaller "effective number" of failures (given by r' and ℓ'), it is optimal.

Theorem 8.4.

Provided $\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \mu$, if $r \geq \ell$ and $r' \leq \ell'$, then $\Delta_n(I) \geq 0$ for all n and I.

Proof:

In view of the conditions, ℓ and ℓ' can be written $r - \delta r$ and $r' + \delta r'$ for $\delta r, \delta r' \geq 0$. Applying first $\hat{J}(n)$ of Theorem 7.2 for $\delta r' = 0$ and then $\hat{K}(n)$ of Theorem 7.2 for $\delta r = 0$,

$$(8.13) \qquad \begin{aligned} \Delta_n(r,r',\mu; L) &\geq \Delta_n(r-\delta r,r',\mu; L) \\ &\geq \Delta_n(r-\delta r,r'+\delta r',\mu; L) \\ &= \Delta_n(\ell,\ell',\mu; L). \end{aligned}$$

By symmetry

$$(8.14) \qquad \Delta_n(\ell,\ell',\mu; L) = \Delta_n(L, L) = 0. \square$$

Theorem 8.4 will be applied in the form of two corollaries; the first gives a sufficient condition for the optimality of $\mathcal{L}$ and the second gives a sufficient condition for the optimality of $\mathcal{R}$, both under the additional condition that $N_{\mathcal{R}} \leq N_{\mathcal{L}}$. They follow immediately from the theorem in view of the logical equivalence of $r \geq \ell$ and $r' \leq \ell'$ when $N_{\mathcal{R}} \leq N_{\mathcal{L}}$.

Corollary 1.

If $N_{\mathcal{R}} \leq N_{\mathcal{L}}$ and $r' \geq \ell'$, then $\Delta_n(I) \leq 0$ for all n and I, provided $\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \mu$.

Corollary 2.

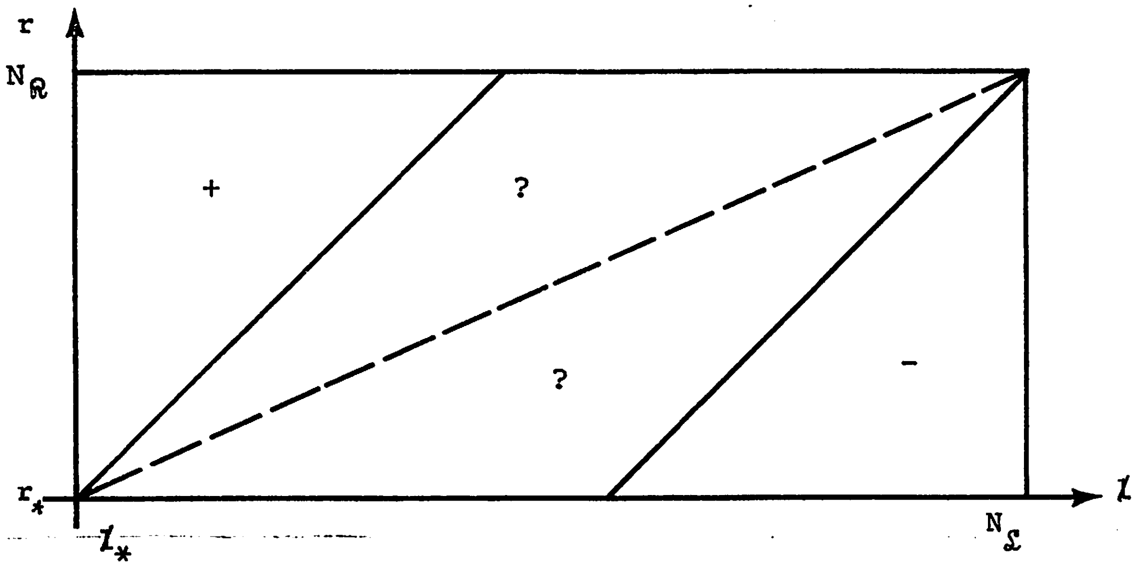
If $N_{\mathcal{R}} \leq N_{\mathcal{L}}$ and $r \geq \ell$, then $\Delta_n(I) \geq 0$ for all n and I, provided $\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \mu$.

If, in addition to the conditions of Theorem 8.4, $n \geq 2$ and $\mu_{\mathfrak{R}}$ is not a one-point measure, then Theorem 7.3 can be applied to strengthen Theorem 8.4. The next theorem is given for completeness, its proof, which will not be given explicitly, uses Theorem 7.3 in the same way that the proof of Theorem 8.4 uses Theorem 7.2.

Theorem 8.5.

Provided $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \mu$ is not a one-point measure, if $r > l$ and $r' \leq l'$ or $r \geq l$ and $r' < l'$, then $\Delta_n(I) > 0$ for $n \geq 2$ and all I.

Figure 8.1 illustrates Corollaries 1 and 2 of Theorem 8.4 for particular values of $N_{\mathfrak{R}}$ and $N_{\mathfrak{L}}$; it is in the style of Figures 3.1 and 3.2, which were for $\mu = \beta$ and particular values of n. The corollaries provide a complete specification of the sign of $\Delta_n(I)$ only if $N_{\mathfrak{R}} = N_{\mathfrak{L}}$, in which case the "?" regions in Figure 8.1 vanish. The figure is somewhat restrictive since $r_* = l_*$ and $r_*' = l_*'$ are implicitly finite, and $N_{\mathfrak{R}} < N_{\mathfrak{L}}$. The dashed line in Figure 8.1 is the locus of points for which $E(\rho|R) = E(\lambda|L)$.



The sign of $\Delta_n(r,r',\mu;\ l,l',\mu)$ for fixed $N_{\mathfrak{R}}$ and $N_{\mathfrak{L}}$ and all n.

FIGURE 8.1

It seems intuitive that Corollary 1 cannot be improved; that is, for given $\mu = \mu_{\mathfrak{R}} = \mu_{\mathcal{L}}$, if $N_{\mathfrak{R}} \leq N_{\mathcal{L}}$ then the only patterns I for which $\Delta_n(I) \leq 0$ for all n have $r' \geq \ell'$. Because, for large n, the effective numbers of successes, r and ℓ, seem to matter less than they do for small n. If not more is known about $\mathfrak{R}$ than about $\mathcal{L}$ (which in a sense is expressed by $N_{\mathfrak{R}} \leq N_{\mathcal{L}}$ whenever $\mu_{\mathfrak{R}} = \mu_{\mathcal{L}}$) and n is large, then obtaining a success on the current pull matters little compared to the possibility of learning something on the current pull about $\mathfrak{R}$ that will increase the number of future pulls on the better arm, except that learning something about arm $\mathfrak{R}$ if arm $\mathcal{L}$ must be used eventually (which is the case if $r' \geq \ell'$) can hardly be very worthwhile. The following conjecture says that Corollary 1 barely holds in the limit as $n \to \infty$; and more, that for a large number of remaining pulls, the only criterion for optimality is the difference between the effective numbers of failures on the two arms.

<u>Conjecture 1.</u>

For any $\mu = \mu_{\mathfrak{R}} = \mu_{\mathcal{L}}$ and all sufficiently large n, $\Delta_n(I)$ has the same sign as $\ell' - r'$, independent of r and ℓ.

On the other hand, Corollary 2 of Theorem 8.4 seems very weak compared to what should be true for all n. For, whenever less is known about arm $\mathfrak{R}$ (and therefore, more information is gained by pulling $\mathfrak{R}$) and $\mathfrak{R}$ offers greater expected immediate payoff, then $\mathfrak{R}$ should be optimal. This is supported by an examination of Figure 3.3; it is noted in Section 3 that for the particular patterns shown, $I = (r, r', \beta; \ell, \ell', \beta)$, where $N_{\mathfrak{R}} = r + r' \leq N_{\mathcal{L}} \doteq \ell + \ell'$, there is no pattern for which $\frac{r}{N_{\mathfrak{R}}} \geq \frac{\ell}{N_{\mathcal{L}}}$ and $\Delta_n(I) < 0$.

Conjecture 2.

For all n and I where $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \mu$, $N_{\mathfrak{R}} \leq N_{\mathfrak{L}}$, and $E(\rho) \geq E(\lambda)$, then $\Delta_n(I) \geq 0$.

Conjecture 2 would imply that for all n, $\Delta_n(I) \geq 0$ for all I in the region above the dashed line in Figure 8.1. Conjecture 2 is implied by the notion that as more becomes known about arm $\mathfrak{R}$, say, and the expected immediate payoff on $\mathfrak{R}$ remains the same ($= E(\rho|R)$), the advantage of $\mathfrak{R}$ over $\mathfrak{L}$ does not increase. This notion can be shown to be equivalent to the next conjecture, which is stated in a manner which emphasizes that it is stronger than $\hat{K}(n)$ of Theorem 7.2.

Conjecture 3.

For all n, for all $I = (r, r', \mu_{\mathfrak{R}}; L)$, and for $\delta r, \delta r' \geq 0$, $\Delta_n(r+\delta r,\ r'+\delta r',\ \mu_{\mathfrak{R}};\ L) \leq \Delta_n(r, r', \mu_{\mathfrak{R}};\ L)$ if $E(\rho|r+\delta r,\ r'+\delta r';\ \mu_{\mathfrak{R}}) \leq E(\rho|r,\ r';\ \mu_{\mathfrak{R}})$.

This conjecture would also imply many instances of the following conjecture in (Chernoff 1968): Let R and L be arbitrary distributions, and R^* a degenerated R, the one-point distribution that concentrates probability one on $E(\rho)$, then $\Delta_n(R, L) \geq 0$ if $\Delta_n(R^*, L) \geq 0$ for all n. This would mean that the solution of the two-armed bandit problem is partially determined by the solution of a corresponding one-armed bandit problem. For any point R in the possibility region of $\mu_{\mathfrak{R}}$, the corresponding R^* is in the direction (a, b) defined by $D_{(a,b)}E(\rho|r,r';\ \mu_{\mathfrak{R}}) = 0$ (provided R^* is in the possibility region for $\mu_{\mathfrak{R}}$; that is, provided R is such that $R(E(\rho) - \varepsilon,\ E(\rho) + \varepsilon) > 0$ for all $\varepsilon > 0$), so that Conjecture 3 would imply

(8.15) $\qquad \Delta_n(R, L) \geq \Delta_n(R^*, L).$

Conjecture 3 would also imply that the solution of the two-armed bandit is partially determined (in the other direction) by the solution of a particular two-armed bandit, one in which one of the arms, say $\mathfrak{R}$, produces either all successes (with probability $E(\rho)$) or all failures (with probability $\bar{E}(\rho)$), and one pull on $\mathfrak{R}$ will, with probability one, reveal which. Let R be an arbitrary distribution with expected value $E(\rho)$, and R_* the distribution which concentrates probabilities $E(\rho)$ and $\bar{E}(\rho)$ at $\rho = 1$ and $\rho = 0$, respectively, then the direction (a, b) in the (r, r') plane from R_* to R (provided R_* is in the possibility region for $\mu_{\mathfrak{R}}$) is defined by $D_{(a,b)}E(\rho|r,r'; \mu_{\mathfrak{R}}) = 0$, and Conjecture 3 would imply

(8.16) $\qquad \Delta_n(R_*, L) \geq \Delta_n(R, L).$

(R_* is in the possibility region for $\mu_{\mathfrak{R}}$ provided R is such that $R[0, \varepsilon) > 0$ and $R(1-\varepsilon, 1] > 0$; if $\mu_{\mathfrak{R}}$ is not such a measure, then (8.16) would follow from Conjecture 3, but for an R_* different from the one defined here.)

9. Results that depend on n

In the previous section, $\hat{J}(n)$ and $J^*(n)$ of Theorems 7.2 and 7.3 are applied for $\delta r' = 0$ and $\hat{K}(n)$ and $K^*(n)$ for $\delta r = 0$, particularly when $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}}$. In the present section, Theorems 7.2 and 7.3 are applied in their full generality when $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}}$. Theorem 8.4 determines the sign of $\Delta_n(I)$ when $r \geq \ell$ and $r' \leq \ell'$ (and, of course, when $r \leq \ell$ and $r' \geq \ell'$); each of the theorems in this section determines the sign of $\Delta_n(I)$ when $r \leq \ell$ and $r' \leq \ell'$ under an additional condition, which depends on n. Theorem 9.1 uses $\hat{J}(n)$ of Theorem 7.2 to determine a sufficient condition for the optimality of $\mathfrak{L}$ and the very closely parallel Theorem 9.2 uses $\hat{K}(n)$ of Theorem 7.2 to determine sufficient conditions for the optimality of $\mathfrak{R}$.

Theorem 9.1.

For all n and $I = (r, r', \mu; \ell, \ell', \mu)$, if $r \leq \ell$ and $r' \leq \ell'$ and $E(\rho | r+n-1, r'; \mu) \leq E(\lambda | \ell+n-1, \ell'; \mu)$, then $\Delta_n(I) \leq 0$.

Proof:

In view of the first two conditions of the theorem, ℓ and ℓ' can be written $r + \delta r$ and $r' + \delta r'$ for $\delta r, \delta r' \geq 0$. The third condition of the theorem then becomes the condition in $\hat{J}(n)$ of Theorem 7.2; therefore,

$$(9.1) \qquad \Delta_n(\ell, \ell', \mu; L) = \Delta_n(r+\delta r, r'+\delta r', \mu; L) \geq \Delta_n(r, r', \mu; L).$$

The conclusion of the theorem follows in view of (8.14). □

Theorem 9.2.

For all n and $I = (r, r', \mu; \ell, \ell', \mu)$, if $r \leq \ell$ and $r' \leq \ell'$ and $E(\rho | r, r'+n-1; \mu) \geq E(\lambda | \ell, \ell'+n-1; \mu)$, then $\Delta_n(I) \geq 0$.

Proof:

The proof of this theorem is strictly parallel to that of Theorem 9.1 with $\hat{K}(n)$ of Theorem 7.2 playing the role of $\hat{\mathfrak{J}}(n)$. □

When μ is not a one-point measure and $n \geq 2$, Theorems 9.1 and 9.2 can be strengthened just as Theorem 8.4 is strengthened by Theorem 8.5. The next two theorems accomplish this. Their proofs will not be given explicitly; they can be proved by applying Theorem 7.3 in the same way that the proofs of Theorems 9.1 and 9.2 apply Theorem 7.2.

Theorem 9.3.

Provided μ is not a one-point measure and $n \geq 2$, if $I = (r,r',\mu;\ \ell,\ell',\mu)$, $r < \ell$ and $r' \leq \ell'$ or $r \leq \ell$ and $r' < \ell'$, and $E(\rho|r+n-1,r';\ \mu) \leq E(\lambda|\ell+n-1,\ell';\ \mu)$, then $\Delta_n(I) < 0$.

Theorem 9.4.

Provided μ is not a one-point measure and $n \geq 2$, if $I = (r,r',\mu;\ \ell,\ell',\mu)$, $r < \ell$ and $r' \leq \ell'$ or $r \leq \ell$ and $r' < \ell'$, and $E(\rho|r,r'+n-1;\ \mu) \geq E(\lambda|\ell,\ell'+n-1;\ \mu)$, then $\Delta_n(I) > 0$.

Theorems 9.1 and 9.2 will be applied to two illustrative example two-armed bandit problems in the following sections: $\mu = \beta$ in the first example and $\mu = \tau$ in the second; β and τ are defined and discussed in Section 2.

10. The beta two-armed bandit

If $\mu = \beta$ as defined by (2.9), the application of Theorems 9.1 and 9.2 is particularly simple. If $R=(r,r'; \beta)$ and $n - 1$ successes are observed in $n - 1$ pulls on $\mathfrak{R}$, the probability of success on the next pull is

$$(10.1) \qquad E(\rho \mid r+n-1, r'; \beta) = \frac{\nu(r,r'; \beta) \int_0^1 \rho^{r+n}(1-\rho)^{r'} d\beta(\rho)}{\nu(r,r'; \beta) \int_0^1 \rho^{r+n-1}(1-\rho)^{r'} d\beta(\rho)}$$

$$= \frac{\int_0^1 \rho^{r+n-1}(1-\rho)^{r'-1} d\rho}{\int_0^1 \rho^{r+n-2}(1-\rho)^{r'-1} d\rho}$$

$$= \frac{r + n - 1}{r + r' + n - 1} = \frac{r + n - 1}{N_{\mathfrak{R}} + n - 1} ;$$

for this formula, see the topic of beta integrals in any advanced calculus text, for example, (Widder 1961, Section 11.2). Similarly, if $R = (r,r'; \beta)$ and $n - 1$ failures are observed in $n - 1$ pulls on $\mathfrak{R}$, the probability of a failure on the next pull (equals one minus the probability of a success) is

$$(10.2) \qquad E(1-\rho \mid r, r'+n-1; \beta) = \frac{r' + n - 1}{N_{\mathfrak{R}} + n - 1} = 1 - \frac{r}{N_{\mathfrak{R}} + n - 1} .$$

Theorems 10.1 and 10.2 apply Theorems 9.1 and 9.2 in a way that complements Theorem 8.4. As in the corollaries of Theorem 8.4, it is assumed for definiteness that the effective number of pulls on $\mathfrak{L}$ is not smaller than the effective number on $\mathfrak{R}$.

Theorem 10.1.

For all n and I, provided $\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \beta$, if $N_{\mathcal{R}} \leq N_{\mathcal{L}}$ and $\frac{r + n - 1}{N_{\mathcal{R}} + n - 1} \leq \frac{\ell + n - 1}{N_{\mathcal{L}} + n - 1}$, then $\Delta_n(I) \leq 0$.

Proof:

First, assume $r' \geq \ell'$. In this case, $r + r' \leq \ell + \ell'$ implies $r \leq \ell$ and, therefore, $\Delta_n(I) \leq 0$ according to Theorem 8.4.

Now, assume $r' < \ell'$. In this case, $\frac{r + n - 1}{N_{\mathcal{R}} + n - 1} \leq \frac{\ell + n - 1}{N_{\mathcal{L}} + n - 1}$ implies $r < \ell$ and, therefore, $\Delta_n(I) \leq 0$ according to Theorem 9.1. □

Theorem 10.2.

For all n and I, provided $\mu_{\mathcal{R}} = \mu_{\mathcal{L}} = \beta$, if $N_{\mathcal{R}} \leq N_{\mathcal{L}}$ and $\frac{r}{N_{\mathcal{R}} + n - 1} \geq \frac{\ell}{N_{\mathcal{L}} + n - 1}$, then $\Delta_n(I) \geq 0$.

Proof:

The proof of this theorem is strictly parallel to that of Theorem 10.1 with Theorem 9.2 playing the role of Theorem 9.1. □

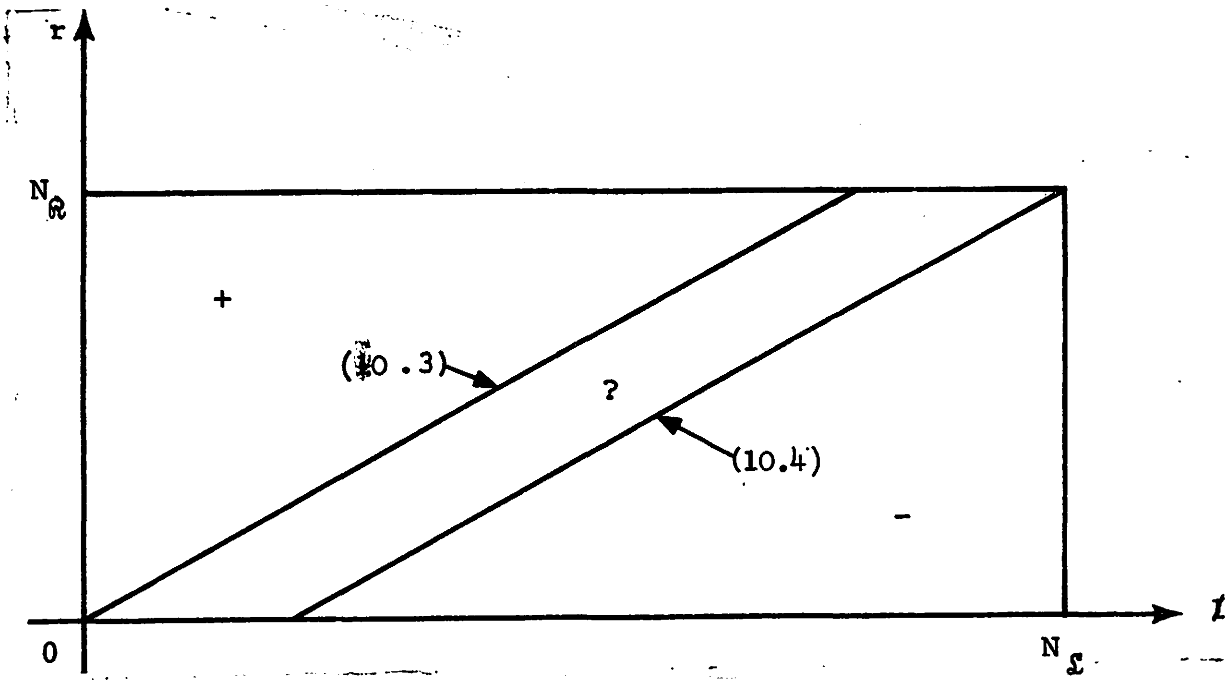
Theorems 10.1 and 10.2 specify the optimal arm for all $I = (r,r',\beta;\ \ell,\ell',\beta)$ only for $n = 1$ (if $N_{\mathcal{R}} \neq N_{\mathcal{L}}$). A typical region in which the sign of $\Delta_n(I)$ is unspecified by these results is shown in Figure 10.1, the equation of the line bounding the "?" region on the left is

$$(10.3)\qquad r = \ell\,\frac{N_{\mathcal{R}} + n - 1}{N_{\mathcal{L}} + n - 1}\,,$$

according to Theorem 10.1, and on the right is

$$(10.4)\qquad r = \ell\,\frac{N_{\mathcal{R}} + n - 1}{N_{\mathcal{L}} + n - 1} - \frac{(N_{\mathcal{L}} - N_{\mathcal{R}})(n-1)}{N_{\mathcal{L}} + n - 1}\,,$$

according to Theorem 10.2.

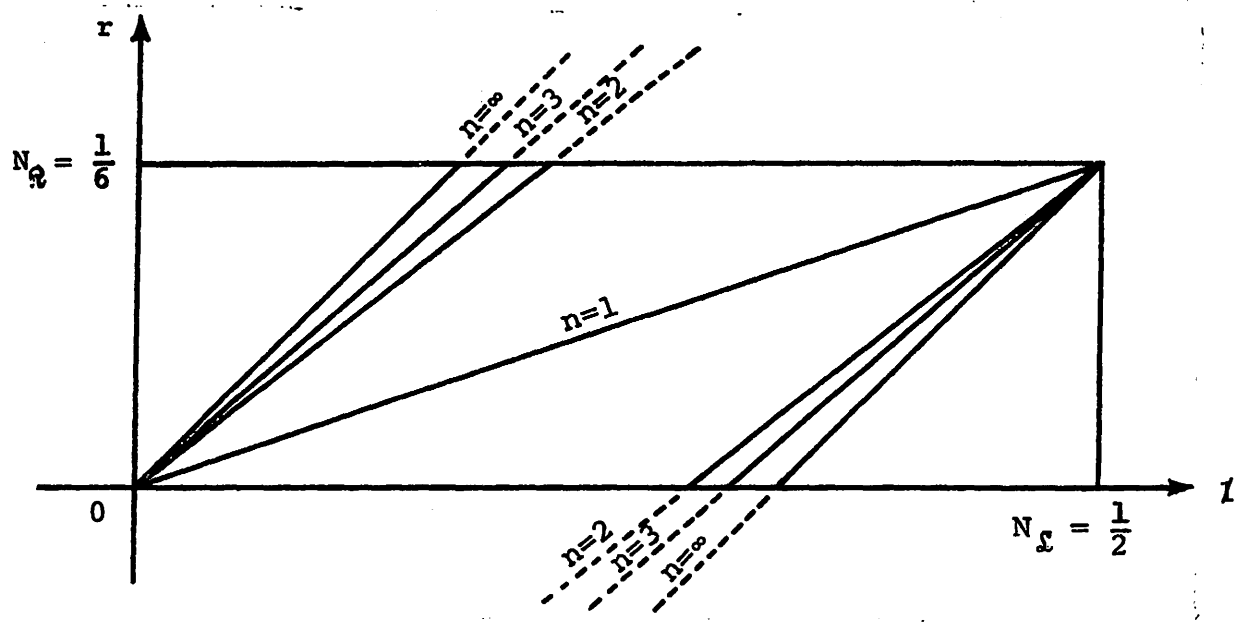


The sign of $\Delta_n(r,r',\beta;\ l,l',\beta)$ for fixed $N_{\mathfrak{R}} = r + r'$ and $N_{\mathfrak{L}} = l + l'$, and a particular value of n.

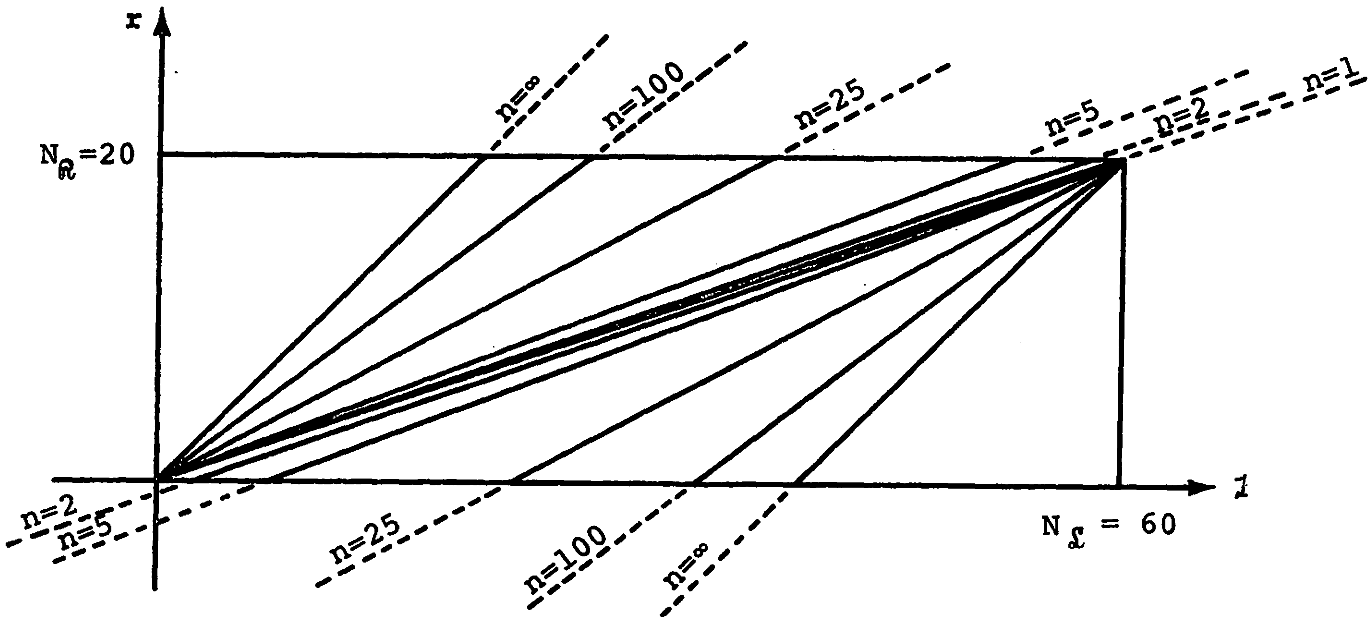
FIGURE 10.1

For $n \to \infty$, the "+" and "-" regions of Figure 10.1 coincide with the "+" and "-" regions of Figure 8.1, where in Figure 8.1, $(r_*, l_*) = (0, 0)$.

While the slope of (10.3) and (10.4) for both extremes, $n = 1$ and $n \to \infty$, is independent of the scale of Figure 10.1, such is not the case for $2 \leq n < \infty$. This effect is illustrated in Figure 10.2, which shows (10.3) and (10.4) for various values of n; in Figure 10.2a, $N_{\mathfrak{R}} = 1/6$ and $N_{\mathfrak{L}} = 1/2$ (and, of course, $0 < r < 1/6$ and $0 < l < 1/2$) while in Figure 10.2b, $N_{\mathfrak{R}} = 20$ and $N_{\mathfrak{L}} = 60$. As is evident, Theorems 10.1 and 10.2 are more helpful for larger values



a. The case: $N_{\mathfrak{R}} = 1/6$, $N_{\mathfrak{L}} = 1/2$.



b. The case: $N_{\mathfrak{R}} = 20$, $N_{\mathfrak{L}} = 60$.

Examples of (10.3) and (10.4) for different values of n, constant $N_{\mathfrak{R}}/N_{\mathfrak{L}}$.

FIGURE 10.2

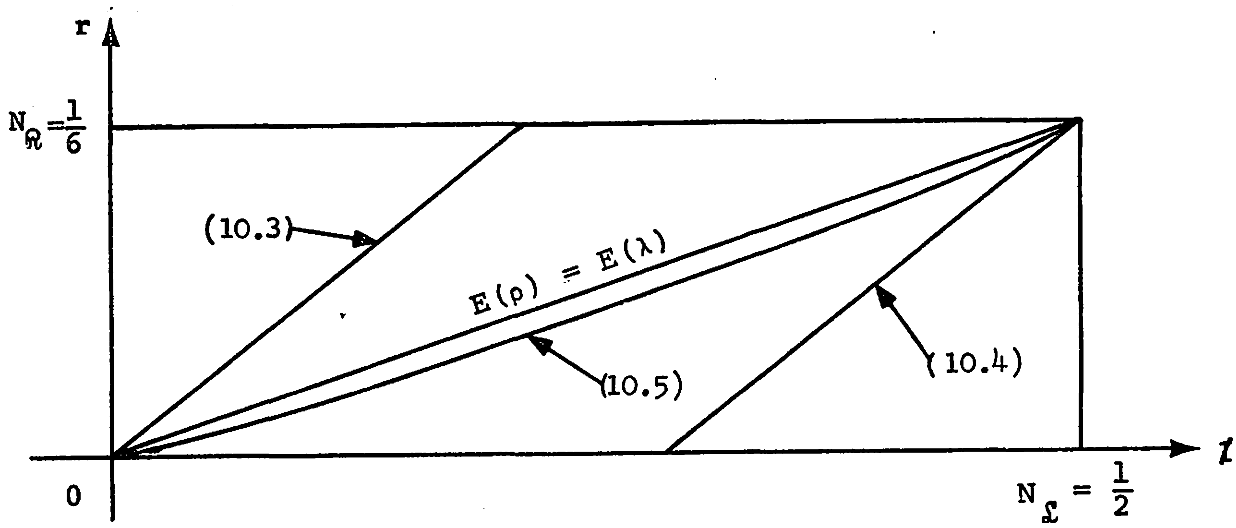
of $N_{\mathfrak{R}}$ and $N_{\mathfrak{L}}$. If $N_{\mathfrak{R}}/N_{\mathfrak{L}}$ is constant while $N_{\mathfrak{R}}, N_{\mathfrak{L}} \to 0$, then for $n > 1$ it provides no improvement on Theorem 8.4.

Theorems 10.1 and 10.2 are compared in Figure 10.3 for $n = 2$ with the exact solution of $\Delta_2(I)$ taken from Section 5 for the case $N_{\mathfrak{R}} = 1/6$, $N_{\mathfrak{L}} = 1/2$. For this case the solution is particularly easy, the only I for which $\mu = \beta$, $N_{\mathfrak{R}} = 1/6$ and $N_{\mathfrak{L}} = 1/2$, and $\Delta_2(I) = 0$ are given by (5.10), case (+0-0), so that $\Delta_2(I)$ has the same sign as

$$E(\rho^2) - E(\lambda^2) = \frac{r}{N_{\mathfrak{R}}}\frac{r+1}{N_{\mathfrak{R}}+1} - \frac{l}{N_{\mathfrak{L}}}\frac{l+1}{N_{\mathfrak{L}}+1} \tag{10.5}$$

$$= \frac{36}{7} r(r+1) - \frac{4}{3} l(l+1).$$

(This represents the difference between the a priori probabilities of two successes on the same arm; i.e., the second moments of R and L.) The curve in Figure 10.3 given by (10.5) specifies those I for which $\Delta_2(I) = 0$. As is evident, neither (10.3) nor (10.4) provide a close approximation to (10.5) [$r = lN_{\mathfrak{R}}/N_{\mathfrak{L}}$ is more reasonable than either!], though (10.4) is closer than (10.3), as suggested by Conjecture 3 at the end of Section 8.



Comparison of $\Delta_2(r,r',\beta;\ l,l',\beta) = 0$, with (10.3) and (10.4) at $n = 2$ for the case: $N_{\mathfrak{R}} = 1/6$, $N_{\mathfrak{L}} = 1/2$.

FIGURE 10.3

11. The two-point two-armed bandit

If $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \tau$, then Theorem 7.2 completely resolves the question of which is the better arm to pull. There is an intuitive reason why this problem is so readily solvable. Ordinarily, $(r, r'; \mu_{\mathfrak{R}})$ is a two-parameter family of distributions. If $\mu_{\mathfrak{R}}$ is not a one-point or two-point measure, then the distribution $(r_1, r_1'; \mu_{\mathfrak{R}})$ is different from the distribution $(r_2, r_2'; \mu_{\mathfrak{R}})$, unless $r_1 = r_2$, $r_1' = r_2'$. But in case $R = (r, r'; \tau)$,

$$(11.1) \qquad \frac{R(\tau_1)}{R(\tau_2)} = \left(\frac{\tau_1}{\tau_2}\right)^r \left(\frac{1 - \tau_1}{1 - \tau_2}\right)^{r'} = \exp\left(r \log \frac{\tau_1}{\tau_2} + r' \log \frac{1 - \tau_1}{1 - \tau_2}\right)$$

$$= \frac{\tau_1}{\tau_2} \exp(r + r'A(\tau)),$$

where $A(\tau)$ is given by (6.15).

In view of (11.1), the whole family of distributions R depends only on the parameter

$$(11.2) \qquad \tilde{r} = r + r'A(\tau).$$

Therefore, for all n, $\Delta_n(r, r', \tau; L)$ depends on (r, r') through $\tilde{r}$ alone and has straight parallel contours in (r, r'). The slope of these contours is $A(\tau)$, the proportion of successes to failures on $\mathfrak{R}$ which does not change Δ_n; for example, if $\tau_1 = 1 - \tau_2$ then $A(\tau) = 1$ and the contours of Δ_n in (r, r') are all parallel to the line $r = r'$.

If the same two numbers τ_1 and τ_2, are the only possible probabilities of success on either arm, then it seems clear that that arm should be pulled which is more likely to be the one associated with τ_2, the larger of the two probabilities; that is, the one which is more likely to be successful on the first pull. The present section will show this to be the case.

Feldman (1962) solved a closed related problem and obtained a similar solution; see also, (Degroot 1970, Section 14.7). In Feldman's problem there are two possible probabilities of success, but the larger is associated with one of the arms and the smaller with the other; which is the better arm is not known. This dependence between the arms is very strong; nevertheless, it will be seen that the solution of Feldman's problem and the solution of the independent two-armed bandit considered in this section can be used to obtain each other.

As previously noted, Theorem 7.2 provides a complete specification of the gradient of $\Delta_n(r,r',\tau; L)$ in (r, r'). Therefore, the sign of $\Delta_n(I)$ is completely determined when $\mu_{\mathfrak{R}}$ and $\mu_{\mathfrak{L}}$ are the same two-point measure. The next theorem shows that the arm which is more likely to yield a success should be pulled.

<u>Theorem 11.1</u>.

For all n and I, provided $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \tau$, $\Delta_n(I)$ has the same sign as $E(\rho) - E(\lambda)$.

<u>Remark</u>.

Theorems 8.4, 9.1, and 9.2 can be cited as in the proofs of Theorems 10.1 and 10.2 to prove Theorem 11.1, However, the fact that $A(r,r'; \tau)$ does not depend on r and r' can be employed more simply to prove the theorem directly by appealing to Theorem 7.2.

Proof of Theorem 11.1:

If $\tau_1 = 0$ and $\tau_2 = 1$ then the theorem is implied by Theorem 4.3(c). Therefore, it will be assumed in the remainder of the proof that either $\tau_1 > 0$ or $\tau_2 < 1$ so that (2.11) can be applied.

According to Theorem 7.2 (see also $J_\tau(n)$ and $K_\tau(n)$ of Theorem 6.1) $\Delta_n(r+\delta r, r'+\delta r', \tau; L) - \Delta_n(r,r',\tau; L)$ has the same sign as $\delta r - \delta r' A(\tau)$; therefore,

$$\Delta_n(r,r',\tau; L) - \Delta_n(\ell,\ell',\tau; L) \tag{11.3}$$

has the same sign as

$$(r-\ell) - (r'-\ell')A(\tau). \tag{11.4}$$

In view of (8.14), where $L = (\ell,\ell'; \tau)$, $\Delta_n(r,r',\tau; L)$ has the same sign as (11.4).

The remainder of the proof will be to show that $E(\rho) - E(\lambda)$ has the same sign as (11.4). According to the assumption that not both $\tau_1 = 0$ and $\tau_2 = 1$, it follows that not both $\tau_1^{\,r}(1-\tau_1)^{r'} = 0$ and $\tau_2^{\,r}(1-\tau_2)^{r'} = 0$. Without loss of generality, assume $\tau_1 > 0$, then in view of (2.11),

$$E(\rho) = \frac{\tau_1 + \tau_2(\frac{\tau_2}{\tau_1})^r(\frac{1-\tau_2}{1-\tau_1})^{r'}}{1 + (\frac{\tau_2}{\tau_1})^r(\frac{1-\tau_2}{1-\tau_1})^{r'}} \tag{11.5}$$

$$= \tau_1 + (\tau_2 - \tau_1)(1 + [(\frac{\tau_2}{\tau_1})^r(\frac{1-\tau_2}{1-\tau_1})^{r'}]^{-1})^{-1} .$$

therefore,

$$(11.6)\qquad E(\rho) - E(\lambda) = (\tau_2 - \tau_1)\{(1 + [(\frac{\tau_2}{\tau_1})^r(\frac{1-\tau_2}{1-\tau_1})^{r'}]^{-1})^{-1} - (1 + [(\frac{\tau_2}{\tau_1})^{\ell}(\frac{1-\tau_2}{1-\tau_1})^{\ell'}]^{-1})^{-1}\}.$$

(11.6) has the same sign as

$$(11.7)\qquad (\frac{\tau_2}{\tau_1})^r(\frac{1-\tau_2}{1-\tau_1})^{r'} - (\frac{\tau_2}{\tau_1})^{\ell}(\frac{1-\tau_2}{1-\tau_1})^{\ell'},$$

and, therefore, the same sign as

$$(11.8)\qquad (r-\ell)\log\frac{\tau_2}{\tau_1} - (r'-\ell')\log\frac{1-\tau_2}{1-\tau_1} = ((r-\ell) - (r'-\ell')A(\tau))\log\frac{\tau_2}{\tau_1}.$$

Since $\tau_2 > \tau_1$, (11.8) has the same sign as (11.4). □

If $\mu_{\mathfrak{R}} = \mu_{\mathfrak{L}} = \tau$ then it is possible that $\rho = \lambda = \tau_1$ (the probability is $R(\tau_1)L(\tau_1)$ since ρ and λ are independent) or that $\rho = \lambda = \tau_2$ (the probability is $R(\tau_2)L(\tau_2)$). If the gambler were given a priori that the arms were identical (that is, either $\rho = \lambda = \tau_1$ or $\rho = \lambda = \tau_2$), then neither arm would be strictly preferred. The only possibilities that influence the size of Δ_n, which determines the preference between the right and left arms, have $\rho \neq \lambda$ (that is, either $\rho = \tau_1, \lambda = \tau_2$ or $\rho = \tau_2, \lambda = \tau_1$). Therefore, $\Delta_n > 0$ when and only when it is a priori more likely that $\rho = \tau_2, \lambda = \tau_1$ than that $\rho = \tau_1, \lambda = \tau_2$.

If the gambler knows a priori that either $\rho = \tau_1, \lambda = \tau_2$ or $\rho = \tau_2, \lambda = \tau_1$, the problem is the same problem considered by Feldman (1962). Therefore, Feldman's result implies and is implied by Theorem 11.1.

References.

Bradt, R. N., S. M. Johnson, and S. Karlin (1956). On Sequential Designs for Maximizing the sum of n Observations. Ann. Math. Stat. 27 1060-1070.

Chernoff, H. (1968). Optimal Stochastic Control. Sankhyā, Series A 30 221-252.

Degroot, M. (1970). Optimal Statistical Decisions. McGraw-Hill, New York.

Dubins, L. E. and L. J. Savage (1965). How to Gamble If You Must: Inequalities for Stochastic Processes. McGraw-Hill, New York.

Feldman, D. (1962). Contributions to the "Two-Armed Bandit" Problem. Ann. Math. Stat. 33 847-856.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II. Wiley, New York.

Hardy, G. H., J. E. Littlewood, and G. Polya (1934). Inequalities. Cambridge University Press, New York and London.

Lehmann, E. L. (1966). Some Concepts of Dependence. Ann. Math. Stat. 37 1137-1153.

Quisel, K. (1965). Extensions of the Two-Armed Bandit and Related Processes with On-Line Experimentation. Stanford Technical Report No. 137, Stanford, California.

Raiffa, H. and R. Schlaifer (1961). Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard University, Boston, Massachusetts.

Widder, D. V. (1961). Advanced Calculus, Second Edition. Prentice-Hall, Englewood Cliffs, New Jersey.

Yakowitz, S. J. (1969). Mathematics of Adaptive Control Processes. American Elsevier Publishing Co., New York.